Gene Expression and CD4+ T Cell Susceptibility to SIV

Gene Expression and CD4+ T Cell Susceptibility to SIV

A dissertation presented

by

Andrew Richard Rahmberg

to

The Division of Medical Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Virology

Harvard University
Cambridge, Massachusetts
August 2016

Dissertation Advisor: Dr. R. Paul Johnson Andrew Richard Rahmberg

Gene Expression and CD4+ T Cell Susceptibility to SIV

Abstract

HIV and its simian counterpart SIV infect CD4+ T cells; however, not all CD4+ T cells are equally vulnerable to infection. Despite decades of research, a complete molecular description of a CD4+ T cell susceptible to infection is still lacking. We hypothesized that the comprehensive expression analysis of a large panel of genes predicted to positively or negatively influence viral replication would provide insights into the molecular mechanisms that underlie differences between subsets of CD4+ T cells in their susceptibility to lentiviral infection. We generated a panel of genes predicted to positively or negatively influence viral replication by mining published data with multiple bioinformatic methods. Using high-throughput microfluidic quantitative real-time PCR, we measured expression of 174 genes in defined subsets of naïve and memory CD4+ T cells from peripheral blood and jejunum. We observed significant modulation of genes in response to memory differentiation, anti-CD3/CD28, and type I interferon stimulation. Despite broad similarities in expression patterns between peripheral blood and jejunum, the highly susceptible CCR5+ transitional memory CD4+ T cells from jejunum had lower total expression of restriction factors relative to the same subset in peripheral blood. Expression of most restriction factors increased during acute SIV infection in all CD4+ T cell subsets.

Gene expression in single cells is known to differ from the average expression obtained by bulk methods. We analyzed over 300 single memory CD4+ T cells from rhesus intestinal tissue during acute SIV infection and identified individual infected and uninfected cells. Expression of 96 genes was compared, revealing PD-1 and CXCR5 as the most differentially

expressed. Together these genes define T follicular helper cells (Tfh); however, Tfh cells have not been previously described in jejunum tissue. We verified the presence of this population and found many phenotypic, transcriptional, and functional similarities of this novel jejunal cell population to classical Tfh cells from lymph nodes. Finally, we found these cells were highly SIV-infected, at much higher levels than total memory CD4+ T cells. Overall, these studies highlighted determinants of susceptibility, identified a novel Tfh-like cell population, and described a prime target for SIV infection during acute infection.

Table of Contents

Acknowledgements

Many people have greatly contributed to my projects, first and foremost to Paul Johnson. I am grateful for his guidance and support while undergoing unforeseen transitions in roles and locations. I also thank present members of the Johnson lab, including Premi Rajakumar for Fluidigm and technical support, Mike Billingsley for statistical and data processing support, Sohni Kaur for flow cytometry, Sama Adnan for valuable discussions, and Barbara Cervasi and Kiran Gill for flow sorting performed at YNPRC. Former lab members with significant contributions include Michelle Connole for flow sorting performed at NEPRC, Jackie Gillis and Arnaud Colantonio for coordination and technical support, and Henoch Hong and Yury Kuzmichev for valuable input in the project's early stages.

I would also like to thank the members of my Dissertation Advisory Committee, Drs. Dana Gabuzda, Todd Allen, and Judy Lieberman for their input over the years, and my thesis examiners, Drs. Alan Engelman, Galit Alter, and Welkin Johnson. Thank you to students in the Virology program, and especially my classmates.

Thank you to my parents and sister for their support throughout my life. Most of all, thank you to my wife Julia da Silva for your unfailing belief in me and constant love. Thank you for following me across the country, and I'm glad we're a team.

List of Figures

List of Tables

Chapter 1: An Introduction to HIV/SIV and the roles of restriction/dependency factors

1.A. Scope of HIV/AIDS global epidemic

Acquired Immune Deficiency Syndrome (AIDS) was initially identified in 1981 as unusual opportunistic infections struck young homosexual men [1]. A lentivirus, now called Human Immunodeficiency Virus 1 (HIV-1), was identified as the causative agent in the following years [2-4]. The pandemic form of HIV-1, called group M, has infected over 65 million people since 1990 and caused over 34 million deaths [5]. Currently, there are over 36 million people living with HIV and over 2 million new infections worldwide in 2015 [6]. While global rates of infection have dropped from their late 1990s peak, areas like Eastern Europe and Central Asia are still seeing rising transmission rates [5]. Even in wealthy countries such as the United States, HIV remains a tremendous burden. Over 1.2 million people are living with HIV infection, and over 650,000 people with an AIDS diagnosis in the US have died overall [7,8].

Antiretroviral drug treatment has had a dramatic impact on disease outcomes. Generally a mixture of three drugs, combination antiretroviral therapy suppresses viral replication reducing plasma viral load to levels undetectable with standard assays (<50 RNA copies/mL) in most patients [9]. This results in significant reconstitution of circulating CD4+ T cells [10], decreases in transmission, decreases in opportunistic infections, and increases in lifespan. Strikingly, a UK cohort with successful treatment resulting in good CD4+ T cell count and low viral load has been found to have normal life expectancy [11].

Despite significant progress in treatment, many challenges still remain. Access to treatment, especially in less wealthy countries, remains challenging. Despite adding about 2 million people to the total currently on treatment in 2014, the rate of new infections still equals or exceeds the rate of new people on treatment [6]. Due to latency and the lifelong nature of infection, patients must continually remain on antiretroviral drugs. Resistance to antiretroviral

drugs is also a significant concern, especially if adherence to proper dosing is not complete. The cost of treatment is also a significant barrier, especially for treatment of HIV-infected people in developing countries.

Over 35 years from the start of the epidemic and despite significant effort, there is still no effective HIV vaccine. Studying the early events of HIV in humans is challenging since individuals may not know they are infected, the acute phase presents with mild symptoms [12], and many of the critical pathogenic events occur in hard to study mucosal tissue [13]. Better understanding of the acute phase of pathogenic lentiviral infections, including characteristics of the early cellular targets of infection, could lead to progress in treatment, vaccine development, and cure research.

1.B. Large-scale screens for host lentiviral dependency factors

Retroviruses have complex lifecycles, yet only encode a small number of proteins and are therefore reliant on many cellular proteins for successful infection. Some host proteins that facilitate different steps of the retroviral lifecycle (termed "dependency factors") have been well studied while others have only been identified in screens with little further study to date. One widely used method to identify host factors required for viral replication is a genome-wide small interfering RNA (siRNA) screen. In general, RNA interference is used to knock down expression of a particular gene, and a resultant phenotype of reduction in viral replication/infectivity signals that the gene was important for the viral replication cycle [14]. Four such studies have been performed for HIV, each using different conditions [15-18]. Brass et al. [17] used HeLa cells expressing CD4 transfected with four siRNAs per gene 72 hours before infection with HIV-1 IIIB. Cells were scored 48 hours later for p24 capsid, and supernatants were tested in new cells

with an integrated tat-responsive β-galactosidase reporter for infectious virion production. Konig et al. [16] used 293T cells, transfected with about 6 siRNAs per gene 48 hours before infection with VSV-G pseudotyped replication defective HIV-1 expressing luciferase. Cells were scored 24 hours later for luciferase and hits were genes that reduced luciferase expression more than 45%. Zhou et al. [15] used HeLa cells expressing CD4 and a β-galactosidase reporter, transfected with about three siRNAs per gene 24 hours before infection with HXB2 replication competent HIV-1. Cells were scored both 48 and 96 hours later for β-gal reporter activation. Finally, Yeung et al. [18] used the Jurkat T cell line, a short-hairpin RNA (shRNA) library with 3-5 shRNAs per gene to stably knock down expression, infection with NL4-3 strain of HIV-1, and survival of cells as a read out.

These four screens alone identified over 1100 genes, and when combined with other smaller studies, nearly 10% of the human genome has been implicated in the HIV lifecycle [19]. There is extremely little overlap between each of the studies, with less than 7% overlap between any pairwise combination of screens [19]. This could be due to a number of reasons, such as different target cells, siRNA libraries, screen conditions, and different filtering of genes whose depletion was toxic. Additionally, not all required genes are likely identified by these screens as they are in cell lines not the natural *in vivo* target cells. Unannotated genes or microRNAs would be missed, and genes with redundant functions would also be overlooked as only one gene is depleted at a time [20]. Many of the identified genes could also be off-target effects, for example if the siRNA depleted multiple genes. To minimize potential off-target effects, one of the screens [17] has been repeated using multiple independent siRNA libraries identifying 108 genes [21]. Nearly 50% of these 108 genes had also been identified in the original screen. Drawbacks aside, large-scale screens have provided valuable insight into host genes required for viral replication.

1.C. Interactions between HIV/SIV and host dependency factors

Binding and Entry

The HIV replication cycle **(Figure 1.1)** begins with adhesion of virus to the host cell, binding of the viral envelope and receptors, fusion of the cell and viral membranes, and delivery of the viral core into the cytoplasm. Initial attachment of the virus to the cell surface can be relatively non-specific and, while non-essential, brings Env into close proximity with viral receptors. Known cellular factors that could contribute include heparin sulfate proteoglycans, which interact non-specifically [22], dendritic cell-specific intercellular adhesion molecular 3-grabbing non-integrin (DC-SIGN) [23], and α4β7 [24,25]. The primary cellular receptor that the Env protein of HIV and SIV bind is the CD4 molecule on the cell surface [26,27]. Binding of the Env trimer causes conformational changes allowing coreceptor binding, mediated in part by the V3 loop [28]. The most common coreceptors for HIV are CCR5 and CXCR4 [29]. SIV in rhesus macaques also primarily uses CCR5 [30], although other coreceptors such as GPR15 and STRL33 can also be used [31]. Coreceptor binding triggers fusion with the host cell membrane by insertion of the gp41 hydrophobic fusion peptide, subsequent six-helix bundle formation, lipid mixing, and formation of the fusion pore [32].

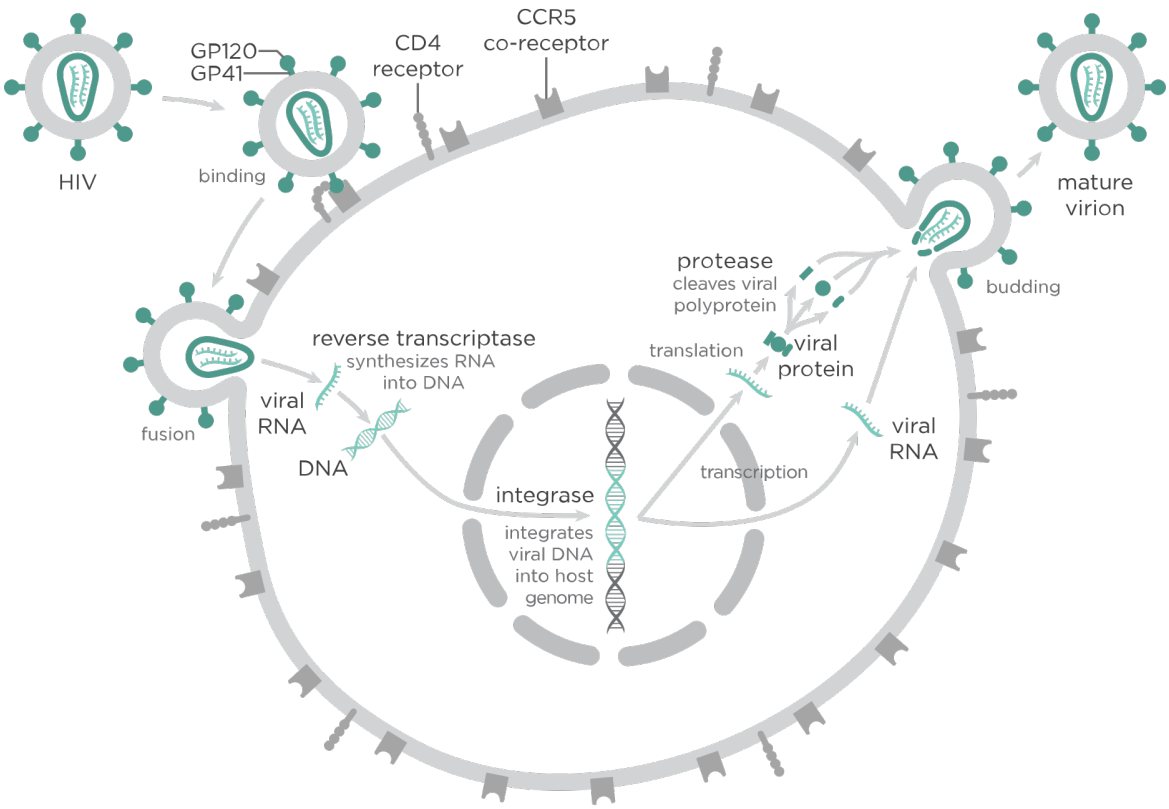


Figure 1.1. HIV replication cycle [33].
The basic steps of an HIV or SIV replication cycle are illustrated including binding and entry, fusion, reverse transcription, integration, production of new viral transcripts and protein, and finally budding and virion maturation.

Reverse Transcription

After fusion, the viral core enters the nucleus, and reverse transcription of the viral genome by the viral enzyme reverse transcriptase (RT) occurs. Although some host factors may play a role in reverse transcription, the polymerase and RNase H activities that are necessary and sufficient are present in RT itself. First, the plus-strand RNA genome is bound by a host tRNA primer near the 5' end of the genome at a site called the primer binding site [34]. In the case of HIV, the host tRNA is Lys3 [35]. RT initiates reverse transcription and generates minus strand DNA, while the RNase H activity degrades the RNA template. The newly synthesized DNA is transferred to the 3' end of the viral RNA and minus strand DNA synthesis continues along the

genome. A polypurine tract is resistant to RNase H cleavage and serves as the primer for plus-strand DNA synthesis [36]. Plus-stand synthesis progresses until the nucleotides of the tRNA are copied allowing for removal of all but one A of the tRNA by RNase H activity [37]. The second (plus-strand) transfer occurs and the ends of both the plus and minus strands are extended to complete the double-stranded linear viral DNA.

Uncoating and Nuclear Import

Reverse transcription in cells occurs in a reverse transcription complex (RTC) containing multiple viral proteins including matrix, capsid, nucleocapsid, integrase (IN), and VPR [38]. Transition of the RTC into a pre-integration complex (PIC) occurs with uncoating, or changes in the structure of the viral capsid core. Uncoating is necessary since HIV infects non-dividing cells and the viral capsid is approximately double the size of the nuclear pore. Many aspects of uncoating including the location, timing, and relationship to reverse transcription are still under investigation [39]. Studies have suggested both removal of capsid prior to nuclear import [40] as well as some capsid remaining associated with the viral complex inside the nucleus [41].

The host protein cyclophilin A (CypA) has long been known to interact with the capsid protein [42,43], specifically a conserved proline-rich loop on the HIV-1 capsid protein [44]. CypA is a host peptidyl prolyl isomerase that can be incorporated into virions [45] but seems to have more of an effect in the cytoplasm of the target cell where it can promote infection in some cell types [46]. Although it seems clear that CypA plays a part in uncoating or in some step in the replication cycle occurring concurrently with uncoating, the current collection of data make it difficult to assign a single critical role [47].

Several cellular factors have been shown to associate with the PIC including HMG I(Y) [48] and LEDGF/p75 [49]. Many nuclear pore components have been found to be associated with

HIV. In fact, three siRNA dependency factor screens were enriched in genes annotated as involved in ‘‘Nuclear Pore / Transport’’ (21–24 genes each), which could participate in processes including the nuclear import of the HIV PIC, export of viral RNAs, or synthesis of other factors [19]. Verified cellular interactors at the level of nuclear import include TNPO3 [50,51], CPSF6 [52], NUP358/RANBP2 [53,54], NUP153 [55] and others.

Integration

Integration is the covalent attachment of the viral DNA to host cell DNA resulting in formation of the provirus and is mediated by the viral IN protein. The catalytic steps of integration are well understood. In general they include 3’ end processing, attack of a pair of phosphodiester bonds separated by five nucleotides in the target DNA, joining of the 3’ ends of viral DNA to the 5’ ends of target DNA, removal of the two unpaired bases at the 5’ ends of the viral DNA, filling in the gaps, and ligation of the 5’ ends of the viral DNA to target DNA [56]. The viral IN protein is responsible for catalyzing the 3’ end processing and DNA strand transfer steps while cellular enzymes are thought to complete the process. *In vitro* integration reactions have been shown to be enhanced by many host factors, including LEDGF/p75, BAF, HMGA, HMGB, Ini-1, and YY1, though the *in vivo* importance of these interactions is unclear [56]. Additionally, several cellular chromatin proteins such as EED, SUV39H1, HP1y, among others have been implicated in integration [57], and cellular DNA repair enzymes likely participate in the final steps of integration, though the important components *in vivo* have not been identified.

The site of integration is not random, with HIV preferring to integrate in active transcription units [58,59]. Viral IN, the chromatin structure, and other genomic features as well as cellular factors help to shape this preference. In particular, the LEDGF/p75 protein has been well

characterized to increase the efficiency of integration and mediate targeting to active transcription units by tethering IN to host chromatin [60-62].

Transcription and Translation

The HIV provirus is a transcription template from which spliced mRNAs for viral protein production and unspliced genomic RNA for packing into new virions is produced. As the virus does not have its own transcription machinery, it relies on the host cell. Although the viral LTR is a promoter; early experiments showed a transactivating factor, called Tat, needed to interact with the transactivation-responsive region (TAR) to allow for efficient transcription [63,64]. Tat is one of the first transcripts produced and works by interacting with the host protein P-TEFb [65-67], which is an elongation factor complex. Tat and P-TEFb work together to remove blocks to elongation by host NELF and DSIF proteins [68,69] while enhancing RNAP II processivity through phosphorylation of its C-terminal domain [70]. In addition to the core promoter region, HIV has an enhancer region with NF-kB binding motifs [71] that members of the NF-kB [72] and NFAT [73] family can bind, supporting replication in primary cells [74]. Multiple subunits of the mediator complex were identified in two or more siRNA screens including MED4, 6, 7, 14, and 28 [19]. This complex links transcriptional activator proteins to the RNA polymerase II transcription apparatus, allowing transcriptional activation. Co-immunoprecipitation suggests an interaction with Tat [75]. Other host proteins with unknown mechanisms of action, such as ZNRD1 [76], also influence transcription.

HIV transcripts undergo splicing using the host splicing machinery (a large complex called a spliceosome) to produce the full range of mRNAs encoding viral proteins. Unspliced and incompletely spliced transcripts, which are normally degraded in the nucleus, are exported

with the help of the viral Rev protein binding the Rev-responsive element in the *env* gene RNA [77,78]. Cellular nuclear import and export pathways are utilized by Rev, and interactions with importin-β, eIF-5A, several FG repeat-containing nucleoporins including Rip/Rab, RanGAP, RanBP1, and CRM1/exportin1 have been identified [79]. Transcripts undergo 3' processing and polyadenylation using many host factors such as CPSF, CstF, CFIm, CFIIm, and poly(A) polymerase [80]. mRNAs are then translated into proteins using normal cellular translational machinery. Some Rev cofactors such as DDX3 have also been implicated in translational regulation [81].

Assembly and Budding

Virion production is accomplished by assembling and packaging essential components, budding where the virion crosses the plasma membrane and acquires its lipid membrane, and maturation where the virion becomes infectious with structural changes. Gag and Gag-Pro-Pol polyproteins are synthesized in the cytosol and traffic to microdomains in the plasma membrane. This requires interaction with many components of intracellular trafficking pathways including AP-1 [82], AP-2 [83], AP-3 [84], KIF4 [85], and ABCE1 [86]. Gag molecules polymerize onto nucleation sites of Gag-RNA complexes, and interactions with matrix promotes Env incorporation. Two copies of noncovalently dimerized, capped, and polyadenylated full-length RNA genome are incorporated into the virion [87]. Gag assembly leads to immature lattice formation, and the virus then uses the host ESCRT pathway to catalyze release [88]. Host proteins TSG101 [89] and ALIX [90] are recruited to sites of budding by Gag late domain motifs, and other components of the host ESCRT machinery such as ESCRT-III (specifically CHMP2 and 4 families) and VSP4 complexes are recruited to complete membrane fission [91]. Concurrently or shortly after budding,

viral maturation occurs by viral protease cleaving Gag and Gag-Pro-Pol polyproteins which then rearrange extensively to create mature infectious virions.

1.D. Interactions between HIV/SIV and host restriction factors

A strict definition of a restriction factor is a host gene that directly and dominantly causes a significant decrease in viral infectivity, has a virally encoded counter-measure, shows signs of rapid evolution, and has expression linked to the host's innate immune response (e.g. is interferon induced) [92]. In a more general sense, a putative restriction factor is any host molecule that can interfere with successful completion of the viral replication cycle, as described here.

Binding and Entry

Blocking the viral envelope protein from interacting with its coreceptors can reduce the ability of virus to enter a cell. Though it is not considered a restriction factor, the level of autocrine CCR5-binding chemokine production (MIP-1α and MIP-1β) has been shown to influence susceptibility to infection [93].

A family of proteins that appear to act at the entry step in HIV infection are the IFITMs [94]. IFITM1, 2, and 3 are ubiquitously expressed, and IFITM1 and 3 respond to interferon stimulation [95]. Though the molecular mechanism behind the IFITM restriction action remains to be elucidated, one suggestion is that IFITM proteins incorporate into the viral membrane and inhibit membrane fusion [96]. Another proposed model is impairing Env protein incorporation in producer cells [97]. Rhesus macaque IFITM proteins are poorly annotated in the existing rhesus genome [98,99]. However, other non-human primate IFITMs show activity against HIV-1 and SIV strains, suggesting a non-species-specific activity [100].

Two recently identified host proteins, SERINC3 and SERINC5, appear to restrict HIV at entry or just post-entry and are counteracted by the Nef protein [101,102]. The studies used independent methods (proteomics and expression profiling) to identify the targets for Nef infectivity enhancement. When viral particles are produced without Nef, large amounts of the SERINCs are incorporated impairing the fusogenicity and potentially early post-entry steps in the replication cycle. When Nef is present in producer cells, SERINCs are redistributed from the plasma membrane into an intracellular Rab7-postive membrane compartment. Despite their apparent antiviral activiety, SERINC3 and 5 do not appear to be under positive selection in primates (including rhesus macaques) unlike confirmed restriction factors APOBEC3G/F, TRIM5α, or BST-2/Tetherin [103]. Nef from SIVmac239 was able to counteract human SERINC5 in human cells [102], though whether it can counter rhesus SERINCs in rhesus cells remains to be investigated.

Reverse Transcription

The host protein SAMHD1 can reduce the efficiency of reverse transcription by reducing the cellular pool of dNTPs with dNTPase activity [104,105]. The lack of sufficient dNTPs leads to a failure to complete reverse transcription quickly enough to avoid particle disintegration and degradation. This restriction is cell type-specific, since myeloid cells [106] and quiescent CD4+ T cells [107,108] have lower steady-state levels of dNTPs than activated replicating cells. The viral protein Vpx, found in many SIV strains and HIV-2 though not HIV-1 [109], counteracts SAMHD1 by targeting it for proteosomal degradation by ubiquitination through the CUL4A/DDB1/DCAF1 pathway [110-112] . The regulation of SAMHD1 activity is still under investigation; however, phosphorylation by cellular CDK proteins [113] and p21 [114], an inhibitor of CDK2, appear to be

involved. As CDK protein expression varies during the cell cycle, other layers of regulation are likely involved as well.

The APOBEC family of restriction factors acts by deaminating C-to-U bases in viral cDNA upon initiation of reverse transcription. This results in G-to-A hypermutation in viral sequences [115]. In cells producing virus, APOBEC3D, F, G, and H can package into virions, and so are poised to function as soon as reverse transcription occurs [116]. All lentiviruses, except equine infectious anemia virus, express a Vif protein optimized to neutralize the restrictive APOBEC3 proteins of their host species [117]. Vif overcomes APOBEC3 restriction by binding host protein CBFβ and recruiting a host E3 ubiquitin ligase complex (RBX2, CUL5, ELOB and C) to target them for degradation by the 26 S proteasome [92,118].

First identified in a large scale RNAi screen for HIV restriction factors [119], RPRD2 / REAF has been found to act at an early stage in the viral life cycle reducing early and late reverse transcripts which subsequently impacts the level of integration [120]. The mechanism of action is still under investigation, although its association with viral nucleic acids suggests it may prevent reverse transcription.

Uncoating and Nuclear Import

A screen for rhesus macaque genes that could restrict HIV-1 infection when expressed in human cells first identified TRIM5α [121]. The mechanism of action is still under investigation, but it is known that TRIM5α binds directly to retroviral capsids [122]. Models have been proposed that suggest an accelerated capsid fragmentation in the cytoplasm soon after entry, which disrupts reverse transcription complexes. However, proteosome inhibition that prevents TRIM5-promoted capsid disassembly restores reverse transcription but not infection [123], indicating this mechanism alone fails to account for the full inhibitory potential of TRIM5α. TRIM5 has also been proposed

to be a pattern recognition receptor for retroviral capsids as it can lead to a signaling cascade activating AP-1 and NF-kB-dependent genes [124]. This could help to establish an antiviral state and account for restriction beyond capsid fragmentation.

TRIM5α is species-specific in its activity, with human TRIM5 being relatively ineffective against HIV-1 (though active against some other retroviruses), while TRIM5α from Old World monkey species generally inhibit HIV [121]. This makes TRIM5α a barrier to cross-species transmission of primate lentiviruses [125,126].

TRIM5 is part of a family of TRIM genes, many of which have been shown to have anti-HIV activity [127] and to be interferon induced [128]. Studies of other TRIM family members have shown inhibition at multiple stages in the HIV replication cycle [129,130].

Recently, the interferon-inducible protein MxB / Mx2, was found to restrict HIV in a capsid-dependent manner [131-133]. Reverse transcription occurred normally in overexpression systems, but integration was reduced. Precise mechanisms for restriction are still under investigation, but it appears that interaction with capsid is required and that dimerization of MxB molecules supports this interaction [134].

Integration

Not all viral DNA entering the nucleus becomes productively integrated, since dead-end pathways of both one- and two-LTR circles can be found. Inactivation of host cell nonhomologous end joining components Ku70/80, ligase IV, and XRCC4 blocks 2-LTR circle formation [135,136], the cellular MRN complex has been implicated in 1-LTR circle formation [137], and DNA repair enzymes XPB and XPD have been reported to inhibit replication [138].

Transcription and Translation

Limited reports suggest the host protein ZC3H12A/MCPIP1 can reduce HIV infection by decreasing steady state levels of the viral mRNA though its RNase domain and that SIV Vpx does not counteract this protein [139,140].

Components of the exosome complex (EXOSC2, 3, and 10) were top hits in a large-scale siRNA screen for HIV restriction factors [119]. RNA exosomes reside in the nucleus and cytoplasm and are involved in RNA quality control and decay, and so could reduce levels of HIV transcripts.

Viral translation can be inhibited by SLFN11 [141]. Wild-type isolates of HIV-1 have a different codon bias than the host cell, and the supply of tRNAs generally increases during HIV infection to account for HIV's requirement for rarer tRNAs [142]. SLFN11 can counteract HIV-induced changes in tRNA levels by a direct interaction with tRNAs, thereby inhibiting viral protein production.

Assembly and Budding

A verified host restriction factor that functions by impairing viral release, BST2 / tetherin, was discovered as the protein that accounts for the late-stage defect in Vpu-deficient HIV-1 particle release [143,144]. Tetherin can incorporate into viral membranes and physically bridge particles to the cell plasma membrane, where captured virions can be internalized for lysosomal degradation [145]. HIV-1's Vpu protein recruits β-TrCP2, a component of the SKP1-CUL-F box E3 ubiquitin ligase complex, leading to the down-regulation and degradation of tetherin [146]. Viruses without *vpu* genes use alternative mechanisms to counteract restriction by tetherin, emphasizing the importance of overcoming this block to viral replication. SIVmac uses Nef [147] and AP-2 dependent endocytosis [148], while HIV-2 uses its Env protein and clatherin dependent internalization and sequestration [149] to counteract tetherin.

1.E. Rhesus macaques and SIV as a model system for humans and HIV

Rhesus macaques (*Macaca mulatta*) and humans diverged evolutionarily about 25 million years ago and share roughly 93% genomic sequence identity [99]. Macaques exhibit greater similarity to humans in physiology, neurobiology, and susceptibility to infectious diseases than do rodents, which are separated from humans by more than 70 million years [99]. Rhesus macaques are the most widely used nonhuman primate species in biomedical research, and a wealth of data exists on their physiology, endocrinology, and metabolism [150]. Importantly, the proportions of the major intestinal T cell subsets in humans and normal macaques are very similar, making macaques a good model for critical phases of immunodeficiency virus pathogenesis [151]. Of potential importance to this study, the similarity of T follicular helper (Tfh) cells appears greater between humans and non-human primates than between humans and mice. For example, the effects of STAT3 and 4 transcription factors on Tfh differentiation differ between humans and mice [152], and the cytokine Activin A was recently shown to drive human and rhesus macaque, but not mouse, Tfh differentiation, despite 100% sequence identity between human and mouse Activin A [153].

Both SIVmac and HIV-1 are lentiviruses encoding most of the same proteins, including Gag, Pol, and Env, as well as accessory proteins Tat, Rev, Nef, Vif, and Vpr. HIV-1 encodes Vpu while SIVmac, like the SIVsm lineage it was derived from, encodes Vpx [154]. Both HIV and SIV use CD4 as a primary entry receptor with CCR5 as the primary coreceptor [29,155]. Their replication cycles are similar, and they use many of the same host proteins to replicate. For example, the preference for integration in actively transcribed genomic regions without

preference for CpG islands or transcription start sites is the same in HIV-1 and SIV [156]. The differences in accessory proteins often stem from adaptation to their host species [109].

HIV-1 originated in humans as cross-species transmission events from chimpanzee SIV [157]. Each HIV-1 lineage, termed groups M, N, O, and P, are derived from an individual transmission event from SIVcpz (groups M and N) or SIVgor [158]. Similarly, SIV in rhesus macaques (SIVmac) arose as a cross-species transmission in United States primate centers from SIV in sooty mangabys (SIVsm) [159]. Natural host species for SIV generally fail to progress to AIDS despite active viral replication [160]. HIV-1 and SIVmac generally do cause AIDS in their hosts, a similarity underlined by their relatively recent cross-species origins.

SIVmac was isolated in rhesus macaques shortly after identification of HIV-1 [161], and the similar disease course was quickly recognized [162]. AIDS in both species results in generalized immune activation, CD4+ T cell depletion (especially from mucosal sites) [163], opportunistic infections, weight loss, and wasting [164]. Critically, rhesus macaques and humans share a hallmark feature of HIV and SIV infection: the dramatic depletion of CD4+ T cells in the gut mucosa [165,166]. The loss of these cells, which are not fully reconstituted even on antiretroviral therapy [167,168], could underlie eventual immune dysfunction.

Advantages of the SIV and rhesus macaque system are many. Animals can be experimentally infected such that the time of infection is precisely known. This allows for study of the earliest phases of infection, at a time before most infected humans are aware of infection. Animals can be sacrificed, allowing access to tissue in locations and in amounts that would be unobtainable in humans. Experimental manipulations that worsen disease such as CD8+ T cell depletion during infection can be tested in macaques [169]. The route and dose of infection can be chosen such that it either approximates most human infections using a repeated low dose

mucosal route [170] or ensures a reliable and synchronous infection by using a high dose intravenously. Additional routes of infection that also mimic human acquisition have been developed [164]. Many years of experience with SIV and rhesus macaques have resulted in well-known viral kinetics such that the peak of acute infection can be relatively accurately predicted [171]. Finally, the viral inoculum is known and a particular sequenced clone can be utilized, simplifying PCR-based viral quantification.

1.F. Pathogenesis and CD4 depletion

The classical phases of HIV infection are viral transmission and acquisition, acute/early infection, viral set point and clinical latency, and AIDS/NeuroAIDS **(Figure 1.1)**. AIDS is a profound immunodeficiency due to loss of CD4+ T cells, classified in part by measuring fewer than 200 CD4+ T cells per μl of blood plasma. On average, about half of untreated HIV-positive persons will develop major opportunistic complications within 10 years of becoming infected [13].

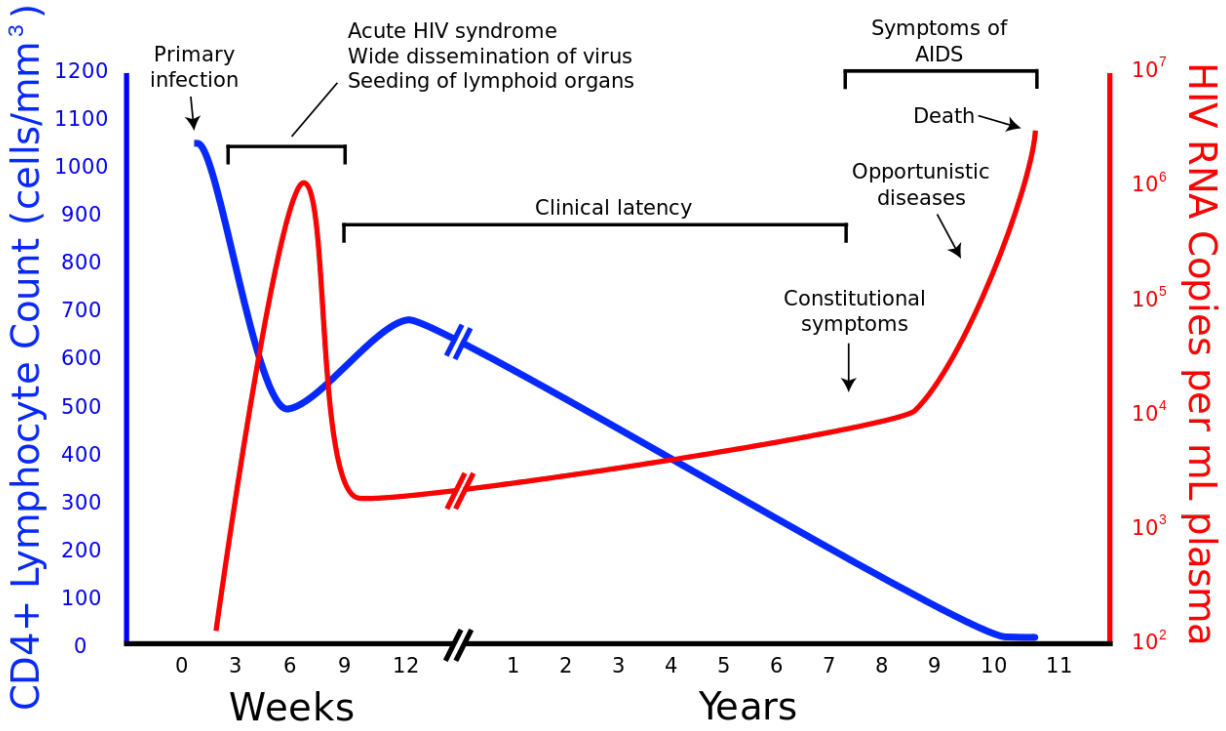


Figure 1.2. HIV disease course [172].
A typical course of HIV infection is displayed comparing plasma viral load and peripheral blood CD4+ T cell count. Experimental infection of rhesus macaques with high doses of SIV will progress faster with a peak of acute infection 10-14 days post infection.

In general, HIV infections are established by a single viral variant in sexual transmission or one to a few variants in intravenous drug use or mother-to-child transmission [173,174]. The majority of transmissions are across a mucosal surface: lower intestinal tract or colorectal mucosa through infected body fluids and semen in sexual transmission, and upper intestinal tract in mother to child transmission from blood/cervical mucus during delivery or from breast milk [175]. There is an initial focus of infection in activated and resting CD4+ T cells in or near the epithelial layer [176,177]. Dendritic cells may facilitate the initial CD4+ T cell infection but their *in vivo* relevance is not established [23,178,179]. Replication then spreads to proximal lymphoid organs by multiple potential mechanisms including free virus diffusion, travel of infected T cells, and/or transport by dendritic cells [13]. During the peak of acute infection, CD4+ T cells are rapidly

depleted from the gut mucosa [163,165,166,180]. This depletion can encompass over one half of the memory CD4+ T cells during the acute phase alone [165]. Increased replication in lymphoid tissue subsequently leads to systemic infection. For example, two weeks after intravenous infection in rhesus macaques, SIV is found at relatively high levels in all lymphoid organs [13]. During the course of chronic infection, inflammation, increased cell turnover, and possibly underlying microbial translocation are factors that contribute to the slow decline of CD4+ T cells in most infected persons to the point at which there is profound immunodeficiency in both T and B cell responses [13,181]. This profound loss of CD4+ T cells results in compromised immune responses to infections and malignancies related to viral pathogens.

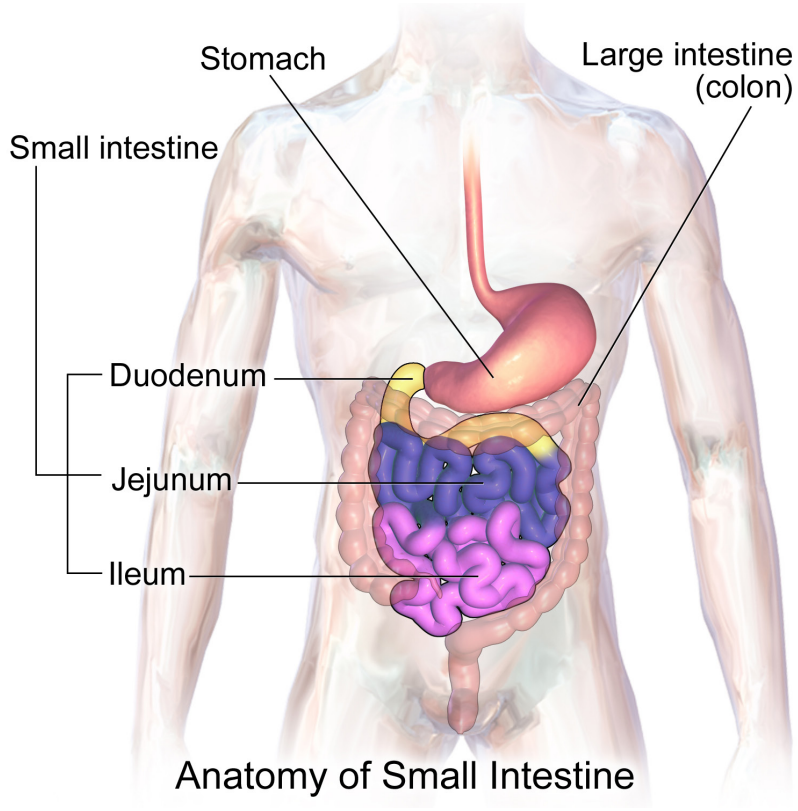


Figure 1.3. Regions of the intestine [182].
Though CD4+ T cells are depleted throughout the intestinal immune system by HIV and SIV, we focused on jejunum tissue since it is more abundant than ileum in rhesus macaques.

Since CD4+ T cell depletion is the primary cause of immunodeficiency, a large body of research has focused on mechanisms of CD4+ T cell loss. The kinetics of CD4+ T cell loss are too rapid to be explained by impairment of cell production, implying direct or indirect mechanisms of cell death. While elimination of infected CD4+ T cells by immune mechanisms like cytotoxic T lymphocytes (CTLs) is certainly possible, depletion of CD8 T cells did not change the lifespan of productively SIV infected cells *in vivo*, indicating a virus-driven mechanism [183]. However, infected monocyte-derived macrophages can produce virus *in vitro* for weeks, whereas infected CD4+ T cells die in days, suggesting viral replication alone does not always result in killing [184]. The short half-life of virus in blood after antiretroviral therapy is initiated suggests CD4+ T cells are killed as rapidly *in vivo* as *in vitro* [185]. Possible viral mechanisms of CD4+ T cell destruction can be categorized as either direct or indirect.

Various viral mechanisms have been implicated for direct viral killing of CD4+ T cells. Envelope protein expression alone can be toxic [186] and surface expression can lead to interaction with nearby cells and formation of syncytia and apoptosis [187]. Membrane disruption due to budding [188] and accumulation of unintegrated DNA [189] have also been posited as direct mechanisms of CD4+ T cell death. The HIV protease can induce apoptosis through a caspase 8/caspase 9 mechanism [190]. Additional mechanisms from other viral proteins have also been suggested [191].

Killing of non-productively infected bystander cells, or indirect killing, is also likely to play a significant role in HIV and SIV pathogenesis. The vast majority of cells undergoing apoptosis in lymph nodes of infected children or rhesus macaques in one study were virus-negative [192]. By *in situ* hybridization for SIV RNA, the total number of cells infected and lost from days 6 to 28 after inoculation could account for only 20% of the total observed CD4+ T cell

depletion [180]. Many potential mechanisms exist for bystander killing, such as activation-induced cell death, immune activation increasing Fas and Fas-ligand expression on T cells, through TNF or TRAIL-mediated pathways, or through soluble viral proteins [191]. More recently, evidence for abortive infection, entry without successful completion of the viral replication cycle, leading to killing of apparent bystander cells through a caspase-3 and -1 mechanism, has been proposed in tonsil tissue, though primarily in the setting of CXCR4-tropic virus [193].

1.G. Immune responses to HIV and SIV

Innate Immunity

The innate immune response is the first line of defense against infection, able to respond in a non-antigen specific manner before adaptive immune responses develop. Cellular components of innate immunity include Langerhans cells, dendritic cells, monocytes, macrophages, natural killer cells, and γδ cells that initially have effector functions [194]. Later in an immune response, innate immune responses may act to support or shape adaptive immunity. Innate immune responses are typically initiated when innate immune cells sense pathogen associated molecular patterns (PAMPs) or damage associated molecular patterns (DAMPs) with pattern recognition receptors (PRRs). The PRRs sensing HIV have long been debated. Proposed PRRs include TRIM5 sensing capsid in T cells [124], cGAS sensing reverse transcription intermediates [195], IFI16 sensing viral DNA in T cells [196], TLR7 sensing viral RNA in plasmacytoid dendritic cells [197], and tetherin in a virion-producing cell [198]. Production of non-cellular components including interferons, chemokines, the complement system, and defensins are consequences of PRR signaling. Restriction factors are also a component of intrinsic innate

immunity since they are capable of acting immediately and their expression is often linked to the innate immune response [92].

Acute HIV infection results in a cytokine storm in which IFN-α, IL-15, TNF-α, and IP-10 are rapidly induced in the first week of infection. Other proinflammatory cytokines/chemokines appear around 7-9 days post-infection, and a more sustained elevation of immunoregulatory cytokines such as IL-10 occurs at later time points [199]. The secretion of MIP-1α (CCL3), MIP-1β (CCL4), and RANTES (CCL5) can prevent SIV infection by binding to and downmodulating the CCR5 coreceptor [200,201]. Interferon production can induce a protective antiviral state, and HIV has long been known to be interferon sensitive both *in vitro* [202] and *in vivo* [203], though a complete understanding of the mechanism of interferon-induced protection is complicated by the large number of induced interferon-stimulated genes.

Innate immune responses to HIV may have both protective and pathogenic consequences. During the initial stages of infection with a small focus of infection, recruitment and activation of additional susceptible cells with proinflammatory signals could allow infection to spread [204-206]. Another clear example is the association between certain NK cell receptors (KIRs) and their corresponding MHC ligands, where certain combinations or copy number variations have been shown to be associated with lower viral loads while others are linked to higher viral loads [207,208].

Macrophages have important functions in tissue immunity and repair, antigen presentation, and tissue homeostasis [209]. While macrophages themselves are not antigen-specific, they can direct epithelial cell maintenance, tissue remodeling, and recruitment of other leukocytes through the production of cytokines and chemokines [209]. However, macrophages are another component of the innate immune system with both protective and pathogenic roles in HIV/SIV infection. Macrophages can contribute to systemic and chronic inflammation, which is

the best predictor of the rate of HIV/SIV disease progression [210]. This inflammatory contribution is thought to occur by multiple mechanisms, including direct infection of macrophages by HIV/SIV, a response to microbial products translocated from the GI tract lumen, phagocytosis of infected CD4+ T cells or antibody complexes, and others.

T cell response

In general, the adaptive immune response to HIV in humans and SIV in macaques is likely a case of too little and/or too late because it fails to prevent systemic spread or to clear the infection [211]. Nevertheless, it is clear that the T cell response does play a role in both acute and chronic infection. Early studies showed cytotoxic lymphocyte-mediated pressure on viral sequences [212] and that the CD8+ T cell response associates with the decline in viremia after acute infection [213,214]. The expression of certain class I HLA alleles strongly associates with disease outcome [215,216], and experimental depletion of CD8+ T cells in macaques demonstrated a key role of this cell population in viral control [169,217]. Additionally, viral evolution at sites of CTL recognition can be detected during peak viremia [218].

Naïve T cells are quiescent, requiring antigen stimulation to become activated [219]. Once activated, CD8+ T cells can function in multiple ways, including direct lysis of virus-infected cells as well as production of cytokines and chemokines such as IFN-γ, IL-2, TNF-α, CCL3, CCL4, CCL5, and others. Debate still exists regarding the relative importance of lytic or cytokine producing mechanisms for control of HIV infection by CD8+ T cells. Rapid viral sequence evolution during acute HIV infection suggests lysis of infected cells is a predominant mechanism [218]. However, depletion of CD8+ T cells after viral set point has been achieved did not result in changes of the lifespan of infected cells, suggesting alternative mechanisms of viral suppression [183,220].

The earliest detectable CD8+ T cell responses to HIV in peripheral blood are narrowly directed to only a few epitopes, predominantly in Env and Nef [218]. These responses are likely functional, since viral evolution at sites of CTL recognition has been detected as early as peak viremia [218]. Interestingly, HIV-specific CD8+ T cell responses in lymph nodes have been shown to be significantly higher than in peripheral blood during chronic infection [221], and very little is known about responses in tissue during acute infection [222]. As infection progresses, the breadth of HIV-specific CD8+ responses increases [223], and the average person targets at least 14 epitopes during chronic infection [224]. A lower viral load is associated with a broader Gag response [225] and polyfunctional T cells exhibiting both cytolytic and cytokine-producing qualities [226]. Elite control, maintenance of a low viral load without treatment, is significantly associated with SNPs in the HLA region [227] and HLA alleles [215,216], emphasizing the importance of the T cell response in chronic infection.

HIV clearly elicits CD4+ T cell responses as well, but these responses generally occur at significantly lower frequencies than CD8+ T cell responses. This is due in part to the dramatic depletion of CD4+ T cells early in infection [228]. HIV-specific CD4+ T cell responses are detected at higher frequencies around the time of peak viremia and are then rapidly lost [229]. CD4+ T cells are important for maintaining CD8 T cell responses. For example, lack of CD4+ T cells in mice significantly impairs CD8+ T cells [230]. Without appropriate CD4+ T cell help, antigen-stimulated CD8+ T cells expand rapidly, become exhausted, and exhibit impaired generation of long-term memory populations, a process requiring IL-2 [231]. The lack or impairment of CD4+ T cell help may be the cause of large magnitude (up to 19% of CD8+ T cells can be HIV-specific [232]) but poorly functional CD8+ T cell response that fails to clear infection.

B cell and antibody response

Similar to the T cell response, the antibody response to HIV is too little and/or too late to prevent systemic establishment or clear infection, and again antibody responses likely play a role in modulating the outcome of infection. Antibodies can function in two general ways: by binding free virions and blocking their binding or entry into cells (neutralization), or by non-neutralizing mechanisms. Non-neutralizing mechanism include virion aggregation; blocking virion transcytosis across epithelial cells; and by binding antigen on an infected cell and interacting with Fc receptors of effector cells, for example phagocytes (Ab-mediated cellular phagocytosis) or natural killer cells (Ab-dependent cellular cytoxicity) [233].

The first detectable B cell response in HIV infection is in the form of immune complexes that develop about a week after detectable viremia [234]. Approximately a week later, free plasma anti-HIV antibody to gp41 appears and is followed about two weeks later by anti-gp120 antibodies [234]. These binding antibodies do not appear to impact the early plasma viral load [234] or exert selective immune pressure [173]. Neutralizing antibodies (NAbs) appear months later, and these can drive neutralization escape [235]. NAbs generally increase in potency over time, and in approximately 30% of individuals antibodies capable of recognizing heterologous virus variants, termed broadly cross neutralizing (BCN) antibodies, eventually develop. Levels of viremia and length of time infected (generally multiple years) seem to be associated with BCN antibody development [236]. Many of the monoclonal BCN antibodies that have been isolated have features such as high levels of hypermutation, suggesting they have undergone multiple rounds of affinity maturation [237]. This observation suggests that, in addition to long-term antigen exposure, functional B cell help in germinal centers where antibody evolution occurs may be critical to effective antibody development.

There is significant interest in antibody-based approaches for HIV vaccine development since protection induced by most vaccines correlates with antibody responses [238,239]. Importantly, pre-existing antibody responses could prevent infection establishment. Additionally, in non-human primates there is significant evidence that passive infusion with high levels of neutralizing antibodies can block infection [240-242] and that more physiologically relevant antibody levels can increase the number of viral challenges needed to infect [243], providing proof-of-principle evidence. Moreover, the only human HIV vaccine trial to date with a signal of efficacy, RV144, appears to have afforded protection by induction of non-neutralizing V2-specific antibodies [244,245]. Finally, as with T cell responses, HIV-specific antibody responses at mucosal sites are poorly understood yet could be critical to effective vaccine protection [246].

1.H. T follicular helper cell biology

T follicular helper (Tfh) cells are a specific subset of CD4+ T cells that is able to provide help to B cells, thereby supporting the induction of antibody responses [152]. Tfh development is a multistage and multisignal process with significant heterogeneity. Accordingly, there is no single event that delineates Tfh differentiation. Multiple redundant signals and pathways for Tfh development are likely to exist, so the following developmental scenario represents a general process that may not apply to all situations. There is an initial priming of a naïve CD4+ T cell by recognition of cognate antigen on a dendritic cell (DC) [247]. Depending on a combination of signals from the DC, including IL-6, ICOS, and the T cell receptor (TCR) strength or dwell time, and the lack of IL-2, the CD4+ T cell undergoes a cell-fate decision within the first few rounds of replication [248-252].

If CXCR5 is expressed, the CD4+ T cell will migrate to the border (interfollicular zone) of the B cell follicle and interact with antigen-specific B cells. This migration is aided by a number of factors: the downregulation of CCR7, a chemotactic receptor for the T cell zone; downregulation of P-selectin glycoprotein ligand 1 (PSGL1); and ICOS-ICOS-ligand binding, (in addition to its function as a costimulatory molecule, can also induce migration of CD4+ T cells) [152,253]. B cells are generally required for development of Tfh cells [254]. B cells serve as the main antigen presenting cells (APCs) in lymph nodes since mature dendritic cells die in a few days during acute infection or immunization. APCs are critical as antigen-specific CD4+ T cells require antigen recognition for each cell division, unlike antigen-specific CD8+ T cells [249].

The third stage of differentiation occurs in the germinal center (GC), a distinct region of the lymph node consisting of stroma, macrophages, follicular dendritic cells, GC B cells, and Tfh cells. Here further interaction with B cells leads to the mature GC Tfh phenotype, which is well conserved across different conditions and species: $CXCR5^{high}$, $PD1^{high}$, $Bcl6^{high}$, Maf^{high}, SAP^{high}, $PSGL1^{low}$, CD200+, $BTLA^{high}$, and $CCR7^{low}$ [152].

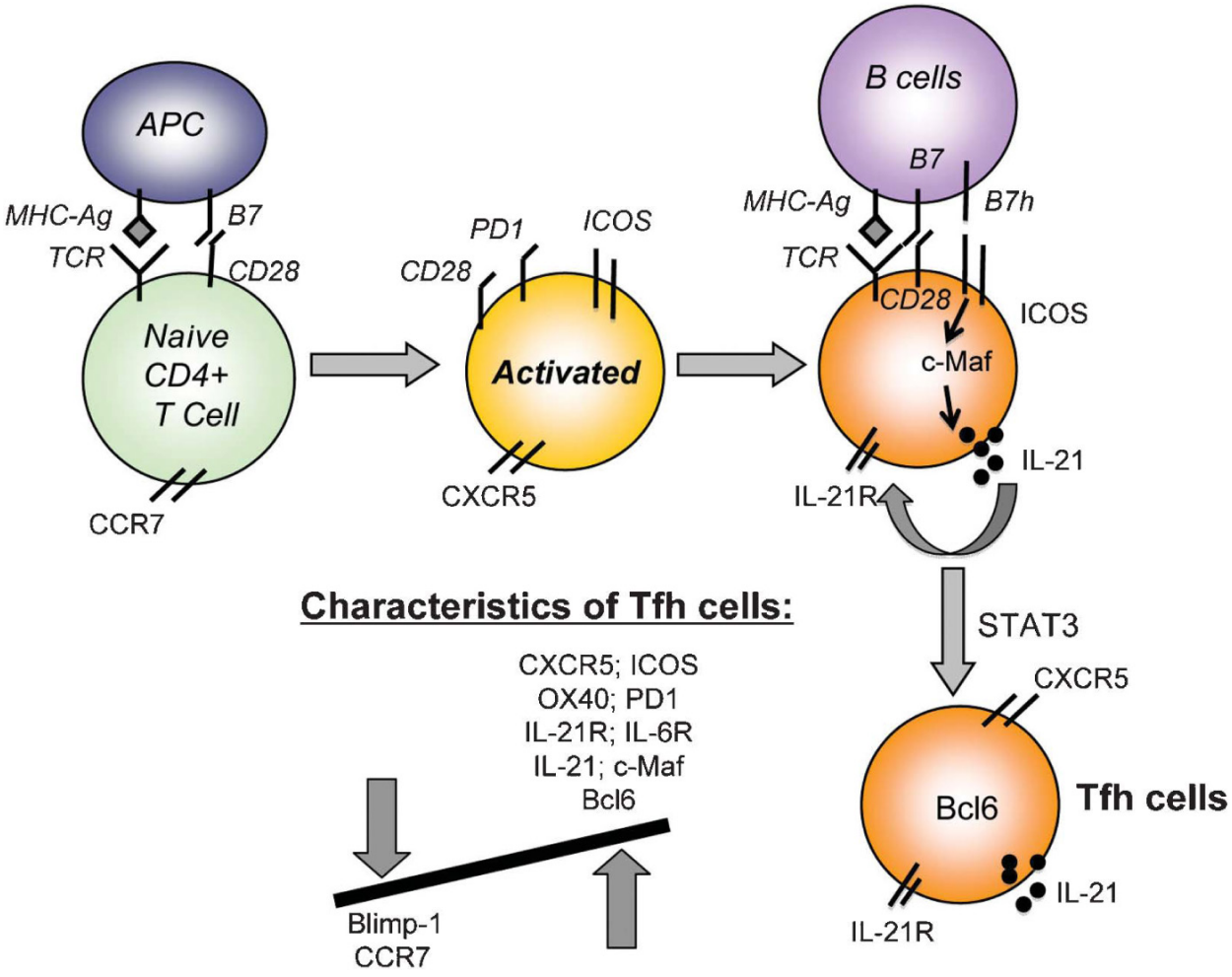


Figure 1.4. Tfh cell generation [255].
Diagrammatic representation of the Tfh development pathway from naïve CD4+ T cells to germinal center Tfh cells. (Reprinted with permission from Nurieva et al., 2010 [255])

Transcription factors important for Tfh differentiation include Maf, IRF4, BATF, STATs, E proteins, and the lineage-defining transcription factor BCL6. BCL6 expression likely promotes Tfh differentiation by multiple mechanisms that have not been fully described. It is known, however, to repress Blimp-1 [256] as well as repress Th1, Th2, Th17, and Treg transcription factors and cytokine genes [251,257].

Tfh cells are required for germinal center formation and function, their primary role in the immune response. Tfh cells regulate GC size [258], support high affinity B cell entry into GCs

[259], and select high affinity B cells [260]. GCs are the main sites for antigen-activated B cell clones to undergo expansion, immunoglobulin gene hypermutation and selection, and affinity maturation. GC B cell selection results in multiple potential outcomes: continued expansion and evolution, apoptosis, or exit from the GC with differentiation into memory B cells or plasma cells [261]. Tfh cells recognizing peptide:MHC complexes on B cells in the light zone of germinal centers provide help signals to GC B cells essential for B cell survival and increased proliferation. GC B cells migrate to the dark zone, proliferate, undergo somatic hypermutation, and migrate back to the light zone to re-present antigen to Tfh cells, at which point the highest affinity B cells (those that bound and endocytosed the most antigen) are selected to undergo another round of proliferation and mutation [152].

The nature of the help signals provided by Tfh cells to GC B cells is still under investigation. CD40L and several canonical secreted molecules of Tfh cells (CXCL13, IL-21, and IL-4) are all involved. Tfh cells are regulated to reduce their own proliferation despite exposure to peptide:MHC complexes. Regulation may be achieved by control of pathways downstream of TCR signaling, high levels of PD-1 expression, and a subset of Tfh cells expressing Foxp3 termed Tfh regulatory cells [262].

Tfh cells have been implicated in multiple aspects of HIV immune response as well as pathogenesis. As their primary role is supporting B cell development and affinity maturation, Tfh cells have been associated with HIV-specific B cell responses, importantly including broadly neutralizing antibodies that generally require extensive hypermutation [263,264]. Tfh cells are also target cells for HIV and SIV infection [265]. It is well established that follicles are major sites of HIV [266-269] and SIV [266,270] replication. One mechanism of enhanced follicular virus replication could be the lack of CD8+ cytotoxic lymphocytes (CTLs) in follicles, as this concentrated

replication does not exist in acute infection prior to the CTL response [270] or after CD8 depletion [271]. Although Tfh cells are depleted in the spleen during acute SIV infection [272], they appear to accumulate in blood [273] and lymph nodes [274] later in infection and could contribute to dysregulated antibody production. Tfh cells have also been suggested as significant part of the latent viral reservoir in both elite controller rhesus macaques and treated aviremic humans [271,275]. Finally, loss of Tfh regulatory cells has been implicated in accumulation of Tfh cells and immune activation in chronic SIV infection [276].

1.I. Single cell gene expression

Widely used methods to measure gene expression such as microarrays and quantitative real-time PCR (qRT-PCR) generally analyze relatively large numbers of cells (10^6 or more). Analysis of gene expression in these studies therefore reflects an average of all the cells in the sampled population. Since cell populations are not homogenous, measurements that average expression over multiple cells can obscure differences among subpopulations. For example, stem cells may make up a vanishingly small proportion of a bulk cell sample yet are critically important to tissue generation [277].

Gene expression on a single cell level is variable such that even housekeeping genes in genetically and cell surface phenotypic marker-identical cells have a broad range of gene expression values [278]. This variability can be due to the intrinsic stochastic nature of biochemical processes regulated by a small number of molecules as well as extrinsic sources such as the cell cycle and the unique signals from the local microenvironment [277,279]. Transcription and translation on the single cell level occurs in periodic bursts [280,281], meaning genes are not expressed uniformly in every cell or at every time. To account for stochastic differences at the

single cell level, many cells and many genes need to be analyzed in order to achieve reliable discrimination of subpopulations. Single cell measurements can deconvolute cell types in putatively homogenous populations, discovering the identity and gene expression profiles of subpopulations, even if they are rare [278].

Early single cell gene expression measurements were limited in the number of cells or genes that could be analyzed [282]. Recent advances in microfluidic devices have allowed significant increases in the throughput in both the number of cells and genes that can be analyzed by single cell qRT-PCR. The Fluidigm C1 Single Cell Auto Prep system allows a suspension of approximately 2000-5000 cells to flow through a microfluidic integrated fluidic circuit (IFC) with 96 single cell capture sites. Once cells are captured and visualized, this system then lyses them, performs reverse transcription, gene-specific pre-amplification, and produces cDNA for each captured cell. Output cDNA is then loaded into a Fluidigm BioMark IFC for qRT-PCR of 96 assays from each captured cell.

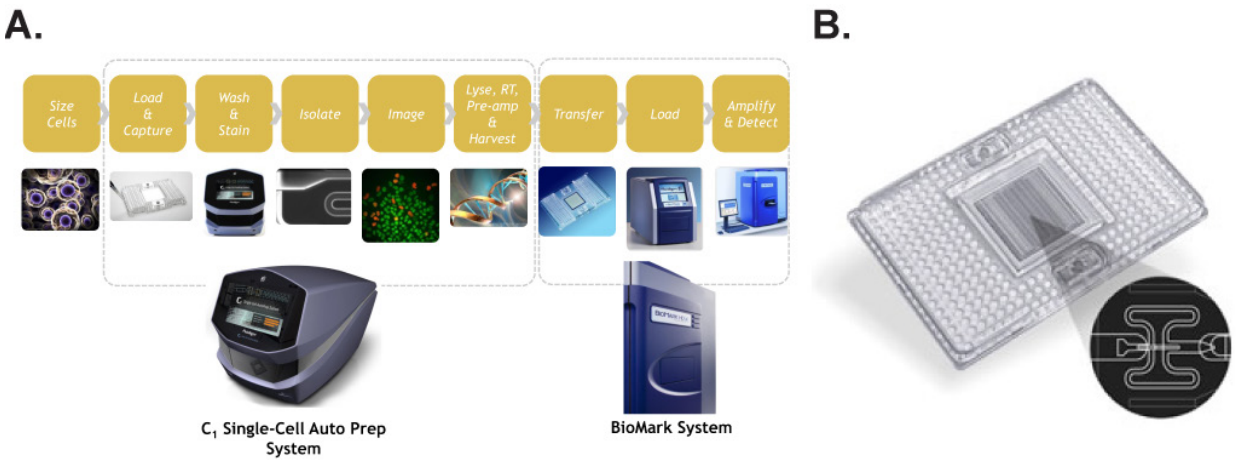


Figure 1.5. Schematic outlining single cell isolation and gene expression analysis procedure. The C1 Single-Cell Auto Prep system is used to perform single cell isolation steps and the BioMark system is used to perform qRT-PCR of up to 96 genes in 96 cells **(A)**. The integrated fluidic circuit (IFC) is a standard sized plate with a microfluidic array at the center **(B)**. The callout shows one of 96 cell capture sites contained in one C1 IFC.

1.J. Scope of the Dissertation

It is known that differences in CD4+ T cell susceptibility to SIV infection exist based on the memory subset [283,284] and the tissue location of the cells [163,165,180]. Very little information exists on the expression of restriction and dependency factor genes in primary cells. Previous studies attempting to comprehensively measure expression of a panel of restriction factor genes [285,286] failed to account for possible differences among memory subsets of CD4+ T cells and only analyzed cells from peripheral blood, not the highly susceptible cells from intestinal mucosa.

In Chapter 2, we hypothesized that comprehensive analysis of expression of confirmed and putative host dependency factors and restriction factors would define a molecular signature of CD4+ T cells most susceptible to infection by SIV. Expression of 45 confirmed and putative restriction factors and 40 dependency factors were analyzed in primary CD4+ T cell memory subsets from peripheral blood and jejunum in both uninfected and acutely SIV infected rhesus macaques.

Gene expression at the single cell level can differ from that seen in bulk measurements of thousands of cells [277,278]. No study has measured single cell gene expression in cells from intestinal CD4+ T cells, with or without viral infection. In Chapter 3, we hypothesized that comprehensive analysis of expression of host dependency factors and restriction factors on a single-cell basis would define a molecular signature of CD4+ T cells susceptible to infection by SIV, revealing a novel subpopulation not identified based on surface phenotype. Over 300 single memory CD4+ T cells from acutely SIV infected rhesus macaque jejunum were analyzed for their expression of 96 genes and SIV transcripts. Individual infected and uninfected cells were identified, and a novel subpopulation of PD-1 and CXCR5 expressing cells were identified as highly SIV susceptible.

Chapter 2: Dynamic modulation of expression of lentiviral restriction factors in primary CD4+ T cells following SIV infection

Acknowledgements

We thank Francois Villinger (Emory University) for providing the $SIV_{MAC}239$ virus stock, Arnaud Colantonio for study coordination, Jackie Gillis and Michelle Connole for flow cytometry support, and Jeffrey Lifson (NCI-Frederick) for providing advice on the SIV PCR primers. We also thank Amber Hoggatt and other members of the NEPRC Division of Veterinary Resources for their expert animal care, and Elizabeth Curran and other members of the NEPRC Pathology Division for their assistance. This study was supported by NIH grants P51 OD011103 and P51 OD011132, as well as funding from the Mucosal Immunology Group (http://public.hivmucosalgroup.org), which is supported by a supplement to the HVTN Laboratory Program (UM1AI068618).

2.A. Abstract

Although multiple restriction factors have been shown to inhibit HIV/SIV replication and multiple genes have been identified as necessary for viral replication, termed dependency factors, little is known about their expression *in vivo*. Expression of 45 confirmed and putative HIV/SIV restriction factors and 40 putative dependency factors was analyzed in CD4+ T cells from peripheral blood and jejunum in rhesus macaques, revealing distinct expression patterns in naïve and memory subsets. In both peripheral blood and jejunum, memory CD4+ T cells expressed higher levels of multiple restriction and dependency factors compared with naïve cells. However, relative to their expression in peripheral blood CD4+ T cells, jejunal CCR5+ CD4+ T cells exhibited significantly lower expression of multiple restriction factors, including *APOBEC3G*, *MX2*, and *TRIM25*, which may contribute to the exquisite susceptibility of these cells to SIV infection. *In vitro* stimulation with anti-CD3/CD28 antibodies or type I interferon resulted in

upregulation of distinct subsets of multiple restriction factors. After infection of rhesus macaques with $SIV_{MAC}239$, expression of most restriction factors substantially increased in all CD4+ T cell memory subsets at the peak of acute infection. Jejunal CCR5+ CD4+ T cells exhibited the highest levels of SIV RNA, corresponding to the lower restriction factor expression in this subset relative to peripheral blood prior to infection. These results illustrate the dynamic modulation of restriction factor expression by memory differentiation, stimulation, tissue microenvironment and SIV infection, and suggest that differential expression of restriction and dependency factors may play a key role in modulating the susceptibility of different populations of CD4+ T cells to lentiviral infection.

2.B. Introduction

Restriction factors serve as a key host defense against virus infection. Many of these genes have well-described activity against the primate lentiviruses HIV and SIV, including the APOBEC3 DNA deaminase family, the TRIM family, *BST-2*/Tetherin, and *SAMHD1* [92]. In addition to the more well-studied restriction factors, screens have been performed to identify additional restriction factors. A whole genome siRNA screen has identified putative restriction factors such as the PAF1 complex and exosome components [119]. A screen for genes sharing genomic characteristics of known restriction factors identified *APOL* and *TNFRSF* family members and used cell-based assays to confirm restriction of HIV-1 [287]. While many studies focus on the impact of a single factor, the total effect of restriction factors on virus infection is likely to be cumulative. Though much work has focused on defining mechanisms of action and structure/function studies for individual restriction factors, little is known about the levels of

expression in primary CD4+ T cells and how expression may be modulated as a result of T cell differentiation and activation or during the course of acute lentiviral infection.

Retroviruses have complex lifecycles, yet only encode a small number of proteins and are therefore reliant on many cellular proteins to complete their replicative cycle. Some host proteins that facilitate different steps of the retroviral lifecycle (termed "dependency factors") have been well-studied while others have only been identified in screens with little further study to date. Similar to restriction factors, a combination of well-studied published genes such as *RANBP2* [54], *NUP153* [55], and putative dependency factors found in screens [15-18,21] have been identified. Also similar to restriction factors, the effect of dependency factor expression on viral replication is likely to be cumulative, as they affect distinct steps in the viral replication cycle.

Naïve CD4+ T cells that are stimulated by cognate antigen can differentiate into a broad range of functionally specialized cell subsets [288]. Studies have found that the differentiation status of a CD4+ T cell influences its susceptibility to HIV and SIV infection, and specifically that memory CD4+ T cells are more likely to be infected than naïve CD4+ T cells [283,284]. The effects of memory differentiation on restriction and dependency factor expression are incompletely understood and may contribute to the differential susceptibility of memory and naïve cells.

During acute infection, HIV and SIV primarily replicate in and deplete gut CD4+ T cells [163,165,180]; however, primary cells from mucosal tissues are relatively understudied as compared to cells from peripheral blood due to the difficulty in obtaining tissue samples [289]. Whether expression of restriction or dependency factors differs between peripheral blood cells, which are infected at lower rates, and cells in the mucosa is currently unknown.

The events of acute immunodeficiency virus infection are challenging to study in humans. Analysis of acute lentiviral infection in nonhuman primates has multiple advantages, including the ability to control the inoculating strain, the precise timing of sampling, and superior access to mucosal and lymphoid tissues. In light of strong evidence documenting the induction of interferon during primary SIV and HIV infection and the fact that many restriction factors are known to be interferon-stimulated genes [206,290], we reasoned that expression of restriction factors is likely to be modulated during the course of SIV infection. However, data on the modulation of expression of restriction factors in different CD4+ T cell subsets during acute SIV infection, especially for the critical CD4+ target cells in the gut mucosa, is not currently defined.

We hypothesized that the comprehensive analysis of expression of a large panel of confirmed and putative restriction and dependency factors would provide insights into the molecular mechanisms that underlie differences between naïve and memory cells in their susceptibility to lentiviral infection as well as the differential infectivity between peripheral blood and gut mucosa CD4+ T cells. Expression analysis of restriction and dependency factors was performed using a high-throughput microfluidic RT-PCR platform that allows for highly quantitative and specific analysis of up to 96 genes at a time from each of 96 samples. Using highly purified sorted populations of CD4+ T cells, we observed both up and down regulation of restriction and dependency factors due to memory differentiation that occurred in a similar pattern in both peripheral blood and jejunum cells. Stimulation with either anti-CD3/CD28 or type I interferon also altered expression of restriction factors in primary cells. Despite broad similarities in expression patterns between peripheral blood and jejunum, the transitional memory CD4+ T cells from jejunum had lower total expression of restriction factors relative to the same subset in peripheral blood. After infection with $SIV_{MAC}239$, the transitional memory

CD4+ T cells from jejunum exhibited the highest level of infection. Strikingly, expression of most restriction factors increased in acute SIV infection in all memory subsets.

2.C. Materials and Methods

Animals

The eight Indian-origin rhesus macaque monkeys (*Macaca mulatta*) described in this study were housed at the New England Primate Research Center (NEPRC) in accordance with the regulations of the American Association of Accreditation of Laboratory Animal Care and the standards of the Association for Assessment and Accreditation of Laboratory Animal Care International. All protocols and procedures were approved by the relevant Institutional Animal Care and Use Committee. Care met the guidance of the Animal Welfare Regulations, OLAW reporting, and the standards set forth in The Guide for the Care and Use of Laboratory Animals [291].

Infection of four animals was performed intravenously with a 500 50% tissue culture infective dose ($TCID_{50}$) of SIVmac239 (generously provided by Francois Villinger (Emory University)). This ensured a reliable and synchronous infection with a well-studied disease course. Euthanasia took place at ten days post-infection using protocols consistent with the American Veterinary Medical Association (AVMA) guidelines.

Lymphocyte isolation

Peripheral blood samples were collected from unvaccinated healthy rhesus macaques for purification of CD4+ T cells. Blood was collected in EDTA vacutainer tubes (Becton Dickinson Vacutainer systems, Franklin Lakes, NJ), and peripheral blood mononuclear cells (PBMC) were

separated by density gradient centrifugation (Lymphocyte Separation Medium; MP Biomedicals Inc., Solon, OH). Jejunum tissue was isolated at time of euthanasia, washed with PBS, separated into small pieces with a scalpel, incubated with 5 mM EDTA for 30 min, washed, and incubated with 15 U/ml type II collagenase (Sigma-Aldrich) followed by mechanical disruption with an 18 gauge-feeding needle and filtration through 70 μm cell strainers (BD Biosciences). Lymphocytes were then enriched by bilayer (35%/60%) isotonic Percoll density gradient centrifugation (1000 × g, 20 min) and the interface containing the lymphocytes was collected [292].

Antibodies and cell sorting

To purify naïve and memory phenotypes, PBMC were stained with CD3 (SP34)-Pacific Blue; CD4 (L200)-FITC; CD8 (RPA-T8)-Alexa 700; CD28 (28.2)-ECD (Beckman Coulter), CD95 (DX2)-APC, CCR5 (3A9)-PE, and CCR7 (2D12)-PE-Cy7. Antibodies were obtained from BD Pharmingen unless specified. PBMCs were initially labeled with LIVE/DEAD viability stain (Life Technologies) and washed, incubated with the CCR7 antibody for 15 min at 37°C, then incubated with all other antibodies for 20 min at room temperature and washed prior to sorting. Cell sorting was performed using a FACS Aria II cell sorter (BD Biosciences). Sorted cell populations were > 99% pure.

In vitro *cell stimulation*

CD4+ T cell enrichment prior to stimulation with αCD3/αCD28 or interferon was performed using the CD4+ T cell non-human primate isolation kit (Miltenyi Biotec). As verified by flow cytometry, purity was > 95%. For anti-CD3/CD28 stimulation, αCD3/αCD28 beads were generated by coupling αCD3 (FN-18) and αCD28 (L293) antibodies with Dynabeads M-

450 Tosylactivated (ThermoFischer Scientific) in a 1:1 ratio as per manufacturer's instructions. Beads were added to cells in a 3:1 ratio for stimulation. For interferon stimulation, recombinant human interferon-αA/D (Sigma-Aldrich) was added to cultures of freshly isolated and enriched CD4+ T cells at a concentration of 1000 units/mL. In both stimulation experiments, paired samples of unstimulated cells were used as controls.

Gene expression analysis

Real-time PCR assays specific for the indicated rhesus macaque genes were purchased from Applied Biosystems (ABI) (Table 1). When not available, macaque-specific custom assays with FAM-MGB probes were designed using Primer Express 3.0 (ABI). Sequences for custom assays were: *APOBEC3D* F-TCCCTGCACTGCAAGCTAAA, R-TGTGTGTGGATACATTGCCTTCA, P-AGATTCTCAGAAACCC; *APOL1* F-CTGGAGGCATCTTGCTTGTG, R-TCTTGCAAGTGCTTTGACTCGTA, P-ATGTGGTCAGCCTTGT; *APOL6* F-AGGCAGAGGAAGAAAGTGAAGCT, R-TCGTCTTCACACAGAGGAACATCT, P-TTGGTTTGGAAAGGGATGAG; *HERC5* F-GGACATACAGATTATGATTGGAAAACA, R-TCACTATGGTGGGATGTGAACTG, P-TTGAAAAGAATGCACGTTATG; *IFITM3A* F-GGCCAGCCTCCCAACTATG, R-GGGCGCCCCCATCAT, P-AAGAGCACGATGTGGC; *IFITM3B* F-AAACCGTCTTCCCTCCTGTCA, R-GCTACCTCATGCTCTTCCTTGAG, P-CCCCCCAGCTATGAG; *SLFN14* F-GAGGGTCTGCAACGACATTTG, R-GCTTCTTACAGAGGGATTCTGGTT, P-TTCCAGTGACACAGCAA; *TRIM14* F-CAGCCAGGAGCCTGATCCT, R-CCGCTGCATCTCCTGCTT, P-AGAGGCTTCAGGCATACA; *TRIM32* F-GGCCTCAATCTGGAGAATCG,

R-AACCAATGGAAAAGCCACCTT, P-CAGAATGAGCACCACCTG.

Dependency factor custom assay sequences were: PIP5K1C F- TGCACATTGGCATCATTGATATC, R- CCAGGTGTGCTCCAGCTTCT, P- CAATCCTACAGGTTCATC; SPG7 F- AGAAGAATAAGGAGAAGGATAAGTCGAA, R- GCCGCTCTCGGTACATCTG, P- ACGAAGAGGATAGGAGACGT; TREX1 F- AGAGTGTGCAGCAAAGTCACTACTG, R- AGCCCATGGTTGGGTAGGA, P- CTGCTACGGCTCAGC; UBASH3A F- GAAGCCAGCAGCAGATGCA, R- CCACCAGCACGCTCTTCCT, P- CCTTGCAGGCCACCG

Primer sequences for the SIV *gag* custom assay were: F-GUCUGCGUCAUCUGGUGCAUUC, R-CAAAACAGAUAGUGCAGAGACACCUAGUG, P- CGCAGAAGAGAAAGUGAAACACACUGAGGAAG.

All assays were confirmed to exhibit linear amplification in an eight point, three-fold dilution series of rhesus PBMC cDNA.

Sorted cells were immediately frozen at -80°C in RNA extraction buffer. RNA extraction was performed on thawed samples using a Qiagen RNeasy Plus Micro kit, RNA purity was confirmed using an Agilent RNA 6000 Pico Bioanalyzer kit, and conversion to cDNA was performed using ABI High Capacity RT with random hexamer primers. cDNA samples were diluted to 200 cell equivalents, linearly pre-amplified with gene-specific primers using an ABI Preamp Master Mix kit and pooled TaqMan assays, and analyzed on a Fluidigm BioMark microfluidic real-time PCR system using 96.96 dynamic arrays. This system allows for the simultaneous measurement of up to 96 assays in 96 different cDNA samples. Initial calculations of cycle thresholds (Ct) were performed using the Fluidigm BioMark software version 4.1.3, and

further analysis was carried out using GenEx software version 6 (MultiD Analyses, URL: http://www.multid.se).

For SIV RT-PCR quantification, an eight-point five-fold standard curve was constructed of *in vitro* transcribed $SIV_{MAC}239$ *gag* RNA and used to interpolate the number of *gag* RNA copies in each cell sample. Copies per cell were calculated based on the number of cells in the reaction.

Statistical analysis

Statistics and graphing were performed using GraphPad Prism version 6.05 for Windows (GraphPad Software, La Jolla California) or R [293]. Naïve and memory CD4+ T cell restriction factor expression was compared using repeated measures ANOVA with Benjamini Hochberg correction for multiple comparisons. All multiple comparison corrections are an extension of the Benjamini Hochberg method [294] updated by Storey [295] and implemented in the R package limma [296]. PCA was performed using R and the function hclust {stats} [293] with resulting axis coordinates and loading factors visualized in GraphPad Prism. Peripheral blood and jejunum memory subsets were compared using two-way ANOVA with Benjamini Hochberg correction. Aggregate peripheral blood and jejunum expression differences were assessed with a paired Student's *t*-test. Pre- and post-infection differences were assessed using a two-way ANOVA with Benjamini Hochberg correction.

2.D. Results

Restriction factor expression in peripheral blood CD4+ T cells

As a comprehensive approach to analyzing the expression of restriction factors in primary CD4+ T cells, a list of confirmed or putative HIV/SIV restriction factors **(Table 2.1)** and dependency factors **(Table 2.2)** was compiled from published sources. To assess expression of restriction and dependency factors in defined cell populations in macaques not infected with SIV, PBMCs and jejunum lymphocytes were isolated from four Indian-origin rhesus macaques, stained with monoclonal antibodies, and sorted into four highly purified populations: naïve (CD3+4+28+95-CCR7+CCR5-), central memory (CD3+4+28+95+CCR7+CCR5-), transitional memory (CD3+4+28+95+CCR7-CCR5+), and effector memory (CD3+4+28-95+CCR7-CCR5-) [297,298]. RNA was extracted and the cDNA samples were analyzed on the Fluidigm BioMark microfluidic real-time PCR system.

Table 2.1. List of restriction factor genes selected for expression analysis.
Assay ID numbers are provided for the ABI TaqMan assays. Sequences for the custom assays are provided in the text.

Gene Name	ENTREZ ID Mm	ENTREZ ID human	HGNC Full Name	Assay ID	Reference
APOBEC3A	702708	200315	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A	Rh04329459_m1	299,300
APOBEC3C	705870	27350	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3C	Rh03418653_s1	299
APOBEC3D	705996	140564	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3D	Custom	116,299
APOBEC3F	723812	200316	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F	Rh04256581_s1	116,299
APOBEC3G	574398	60489	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	Rh02788475_m1	115,116, 299
APOBEC3H	723811	164668	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3H	Rh04256315_m1	116,299
APOL1	694508	8542	apolipoprotein L, 1; APOL; APO-L; FSGS4; APOL-I	Custom	95,301
APOL6	693593	80830	apolipoprotein L, 6	Custom	95,287 302
BST2	719092	684	CD317; NPC-A-7; bone marrow stromal cell antigen 2; Tetherin	Rh02848328_m1	143,144
CD164	699242	8763	CD164 molecule, sialomucin	Rh02859344_m1	287
CDKN1A	719199	1026	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	Hs00355782_m1	302,303

Table 2.1. (Continued)

Gene Name	ENTREZ ID Mm	ENTREZ ID human	HGNC Full Name	Assay ID	Reference
CTR9	705748	9646	Ctr9, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)	Rh02792191_m1	119
EXOSC10	714153	5394	exosome component 10	Rh00897424_m1	119
EXOSC2	715960	23404	exosome component 2	Rh02930061_mH	119
EXOSC3	716347	51010	exosome component 3	Rh02830694_m1	119
GSN	699705	2934	gelsolin (amyloidosis, Finnish type)	Rh02794823_m1	304
HERC5	702743	51191	hect domain and RLD 5	Custom	305
IFITM1	697687	8519	interferon induced transmembrane protein 1 (9-27)	Rh02809735_gH	94
IFITM3A	697829	10410	[predicted] interferon induced transmembrane protein 3-like	Custom	94
IFITM3B	697564	10410	[predicted] interferon induced transmembrane protein 3-like	Custom	94
ISG15	700141	9636	ISG15 ubiquitin-like modifier	Rh02915441_g1	305
MICB	715141	4277	MHC class I polypeptide-related sequence B	Rh02787686_m1	95
MOV10	705910	4343	Mov10, Moloney leukemia virus 10, homolog (mouse)	Rh02878489_m1	306-308
MX2	780935	4600	myxovirus (influenza virus) resistance 2 (mouse)	Rh02801425_m1	131,133
PAF1	697786	54623	Paf1, RNA polymerase II associated factor, homolog (S. cerevisiae)	Rh02876261_m1	119
PARP1	698806	142	poly (ADP-ribose) polymerase 1	Rh00911359_m1	309
PRMT6	694284	55170	protein arginine methyltransferase 6	Rh02817860_s1	310,311
RPRD2 - REAF	715526	23248	regulation of nuclear pre-mRNA domain containing 2; REAF	Rh02868968_m1	119,120
RTF1	706200	23168	Rtf1, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)	Rh01025583_m1	119
SAMHD1	709060	25939	SAM domain and HD domain 1	Rh02869977_m1	104,107
SETDB1	716141	9869	SET domain, bifurcated 1	Rh02803155_m1	119
SLFN11	715511	91607	schlafen family member 11	Rh02885088_m1	141
SLFN14	718850	342618	schlafen family member 14	Custom	
TNFRSF10A	716826	8797	tumor necrosis factor receptor superfamily, member 10a	Rh02846752_m1	287
TNFRSF10D	8793	8793	tumor necrosis factor receptor superfamily, member 10d, decoy with truncated death domain	Rh02846723_m1	287
TRIM14	715418	9830	tripartite motif-containing 14	Custom	95
TRIM19 - PML	700379	5371	promyelocytic leukemia	Rh03043124_m1	312
TRIM22	713814	10346	tripartite motif-containing 22	Rh02801450_m1	129,313, 314
TRIM25	712588	7706	tripartite motif-containing 25	Rh02856605_m1	119
TRIM26	100141397	7726	tripartite motif-containing 26	Rh03418272_m1	127
TRIM28	711982	10155	tripartite motif-containing 28	Rh01076235_m1	130
TRIM32	702595	22954	tripartite motif-containing 32	Custom	315
TRIM34	100568287	53840	TRIM6-TRIM34 readthrough transcript; tripartite motif-containing 6; tripartite motif-containing 34	Rh04256228_m1	316

Table 2.1. (Continued)

TRIM38	694861	10475	tripartite motif-containing 38	Rh02860500_m1	[127]
TRIM5_3_4	574288	85363	tripartite motif containing 5; exon 3-4 junction, all isoforms	Rh02788627_m1	[121,317]
TRIM5A_7_8	574288	85363	tripartite motif containing 5; exon 7-8 junction, alpha isoform only	Rh02788631_m1	[121,317]
ZC3H12A	713604	80149	zinc finger CCCH-type containing 12A; MCPIP1	Rh02882632_mH	[139]

Table 2.2. List of dependency factor genes selected for expression analysis.
Assay ID numbers are provided for the ABI TaqMan assays. Sequences for the custom assays are provided in the text.

Gene Name	ENTREZ ID Mm	ENTREZ ID human	HGNC Full Name	Assay ID
AKT1	697747	207	AKT serine/threonine kinase 1	Rh02834331_m1
CCNT1	707754	904	cyclin T1	Rh02837946_m1
CLDND1	698472	56650	claudin domain containing 1	Rh01026490_m1
CLK2	717671	1196	CDC like kinase 2	Rh02805929_m1
CPSF6	718331	11052	cleavage and polyadenylation specific factor 6	Rh01101216_m1
DDX23	708185	9416	DEAD-box helicase 23	Rh01058445_g1
DMXL1	698117	1657	Dmx like 1	Rh02857893_m1
DNAJB2	702936	3300	DnaJ heat shock protein family (Hsp40) member B2	Rh02800276_m1
EPAS1	714641	2034	endothelial PAS domain protein 1	Rh02838664_m1
FNTA	712681	2339	farnesyltransferase, CAAX box, alpha	Hs00954344_g1
FYB / ADAP	693951	2533	FYN binding protein	Rh02838822_m1
GLUT1 / SLC2A1	698931	6513	solute carrier family 2 member 1	Rh02861051_m1
KLF2	719492	10365	Kruppel like factor 2	Rh02927457_mH
MAP4 NM	100423619	4134	microtubule associated protein 4; one of two possible Macaque homologs	Rh02841978_m1
MAP4 XM	711097	4134	microtubule associated protein 4; one of two possible Macaque homologs	Rh02841981_m1
MED14	699395	9282	mediator complex subunit 14	Rh02847948_m1
MED4	704644	29079	mediator complex subunit 4	Rh02800666_m1
MED6	693320	10001	mediator complex subunit 6	Rh02795099_m1
NUP153	703888	9972	nucleoporin 153	Rh02856816_m1
NUP155	706019	9631	nucleoporin 155	Rh00897468_m1
NUP160	710968	23279	nucleoporin 160	Hs01013191_m1
NUP85	705628	79902	nucleoporin 85	Rh02882139_m1
PIP5K1C	721744	23396	phosphatidylinositol-4-phosphate 5-kinase type 1 gamma	Custom
PSIP1 / LEDGF / P75	664733	11168	PC4 and SFRS1 interacting protein 1	Rh02851085_m1
RAD23A	717979	5886	RAD23 homolog A, nucleotide excision repair protein	Rh02803250_m1
RANBP1	719448	5902	RAN binding protein 1	Rh00862133_g1
RANBP2	694823	5903	RAN binding protein 2	Hs01108576_m1
RAP1B	718171	5908	RAP1B, member of RAS oncogene family	Rh02809565_g1
RBM10	711791	8241	RNA binding motif protein 10	Rh01016429_m1
RELA	715919	5970	RELA proto-oncogene, NF-kB subunit	Rh02879402_m1
SLC38A6	704017	145389	solute carrier family 38 member 6	Hs01553968_m1
SPG7	699488	6687	SPG7, paraplegin matrix AAA peptidase subunit	Custom
STK39	704855	27347	serine/threonine kinase 39	Rh01085344_m1
SUN2	704914	25777	Sad1 and UNC84 domain containing 2	Hs00992164_m1
THOC2	697668	57187	THO complex 2	Rh01107189_m1
TNPO3	700852	23534	transportin 3	Hs01110591_m1

Table 2.2. (Continued)

Gene Name	ENTREZ ID Mm	ENTREZ ID human	HGNC Full Name	Assay ID
TREX1	100429168	11277	three prime repair exonuclease 1	Custom
UBASH3A	722456	53347	ubiquitin associated and SH3 domain containing A	Custom
WNK1	706493	65125	WNK lysine deficient protein kinase 1	Rh01013234_m1
ZNRD1	712151	30834	zinc ribbon domain containing 1	Rh01548529_m1

Sorting for purified populations was critical for isolating the influence of tissue location, since peripheral blood and mucosal tissue differ in the proportions of memory CD4+ T cell populations they harbor **(Table 2.3)**. For example, jejunum CD4+ T cells were comprised of approximately 41% CCR5+ transitional memory (TM) T cells, while peripheral blood CD4+ T cells were only 9% TM phenotype.

Table 2.3. Frequency of CD4+ T cell memory populations in jejunum and peripheral blood.

Tissue	Population	Average Frequency	Minimum Frequency	Maximum Frequency	Number of Animals
Jejunum	Naïve	24	22	29	4
Jejunum	Central Memory	26	29	47	4
Jejunum	Transitional Memory	41	43	66	4
Jejunum	Effector Memory	2	0	4	4
PBMC	Naïve	49	22	69	20
PBMC	Central Memory	31	70	86	20
PBMC	Transitional Memory	9	12	28	20
PBMC	Effector Memory	9	1	20	20

Gene expression results from each animal were normalized by the ΔΔCt method [318] to both the most stable of seven endogenous control genes (*POLR2A*) as determined by the NormFinder algorithm [319] and to the naïve population **(Figure 2.1.A)**. In general, increased expression was seen for many of the restriction factors in memory subsets relative to naïve cells, including *APOBEC* family members, *p21*, *GSN*, and *SAMHD1*. The CCR5+ transitional memory CD4+ T cells from peripheral blood expressed the largest number of restriction factors, 12, with a mean of at least four-fold greater expression relative to naïve cells. By comparison, nine

restriction factors were four-fold greater expressed in effector memory and six in central memory.

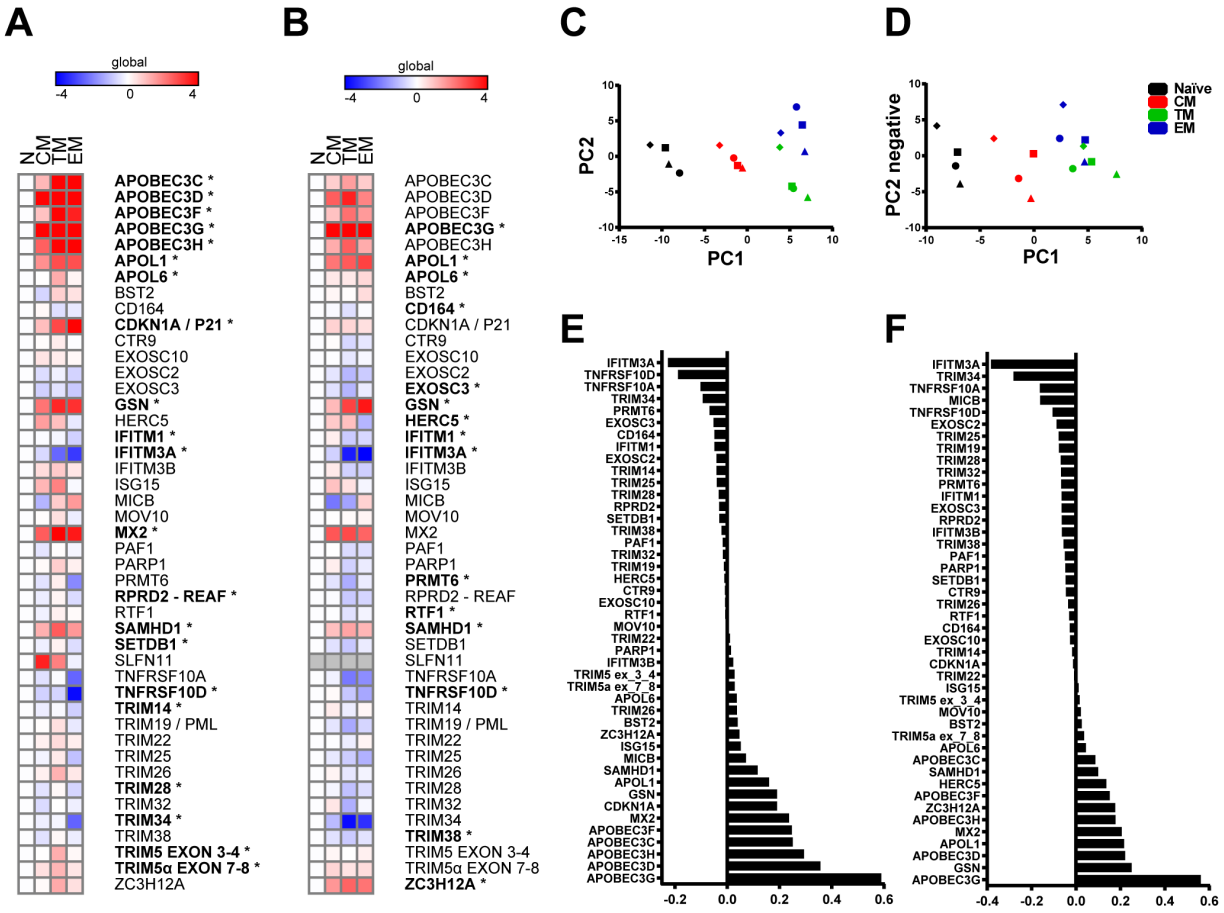


Figure 2.1. Expression of restriction factors exhibits similar patterns in peripheral blood and jejunum CD4+ T cell memory subsets, and is dynamically modulated by stimulation. Heatmap of -ΔΔCt expression values depicting memory subset expression relative to naïve for peripheral blood **(A)** and jejunum **(B)**. Means of four animals, * = $p<0.01$, repeated measures ANOVA with Benjamini Hochberg correction. Principal component analysis for peripheral blood **(C)** and jejunum **(D)** based on analysis of 45 restriction factor genes. Each symbol in (C) and (D) represents a different animal. PC1 loading factors from the PCA for peripheral blood **(E)** and jejunum **(F)**.

Using expression data of only the restriction factor genes normalized to a stable control gene (*POLR2A*), principal component analysis (PCA) revealed expression patterns that clustered by memory subset **(Figure 2.1.C)**. Since each memory subset was more similar to the respective subset from another animal rather than other subsets from the same animal, these data reveal a

reproducible pattern of expression due to memory differentiation. Examination of the loading factors for principal component axis 1 (PC1) identified the genes driving the clustering along the x-axis **(Figure 2.1.E)**. *MX2*, *GSN*, *APOBEC* family members, and *TNFSRF* family members were the major genes contributing to memory subset differences. These genes were among the most differentially expressed restriction factors among memory populations.

Restriction factor expression in jejunum CD4+ T cells

CD4+ T cells from jejunum showed a similar pattern of restriction factor expression as observed in peripheral blood CD4+ T cells, demonstrating that in general, restriction factor expression was modulated primarily by memory differentiation regardless of tissue location **(Figure 2.1.B)**. Similar to peripheral blood, transitional memory cells had the largest number of upregulated restriction factors—eight, relative to naïve cells, as compared to four in central memory and five in effector memory. As in peripheral blood, PCA also clustered memory subsets together for jejunum CD4+ T cells, though transitional and effector memory cells were not as distinctly separated **(Figure 2.1.D)**. Similar genes contributed to the spatial organization of samples along principal component axis 1 (PC1), including *APOBEC*s, *MX2*, *GSN*, and *TNFRSF* genes **(Figure 2.1.F)**.

As *TRIM5* is known to be expressed as multiple differentially spliced isoforms where only the α isoform is restrictive [317], we used two assays for *TRIM5*: one that recognizes a splice junction present in all *TRIM5* transcripts (exon 3-4) and one that recognizes the exon 7-8 splice junction found only in *TRIM5α*. This allowed assessment of the precise functional isoform in case it differed in expression under the various tested conditions. In both peripheral blood and jejunum both *TRIM5* assays showed similar patterns of expression change relative to naïve

CD4+ T cells, indicating the transcript levels are not differentially affected by memory differentiation.

Dependency factor expression in peripheral blood and jejunum CD4+ T cells

As with restriction factors, expression of dependency factors in each memory subset was normalized to naïve cells and the stable control gene *POLR2A* using the ΔΔCt method for both peripheral blood **(Figure 2.2.A)** and jejunum **(Figure 2.2.B)**. An average increase of expression in both tissue compartments was observed. Compared to restriction factors, fewer dependency factors exhibited significant expression modulation due to memory differentiation (17.5% in peripheral blood and 7.5% in jejunum compared to 46.7% and 33.3% respectively for restriction factors). Despite less modulation, PCA still clustered cells from the same naïve or memory subset tighter than cells from the same animal **(Figure 2.2.D** and **Figure 2.2.E)**, demonstrating a reproducible pattern due to memory differentiation. The difference in extent of modulation is apparent when the PCA generated with restriction factors is overlaid on the PCA generated with dependency factors for peripheral blood **(Figure 2.2.F)** and jejunum **(Figure 2.2.G)**. The wider spread of restriction factor subsets is reflective of the larger extent of expression modulation.

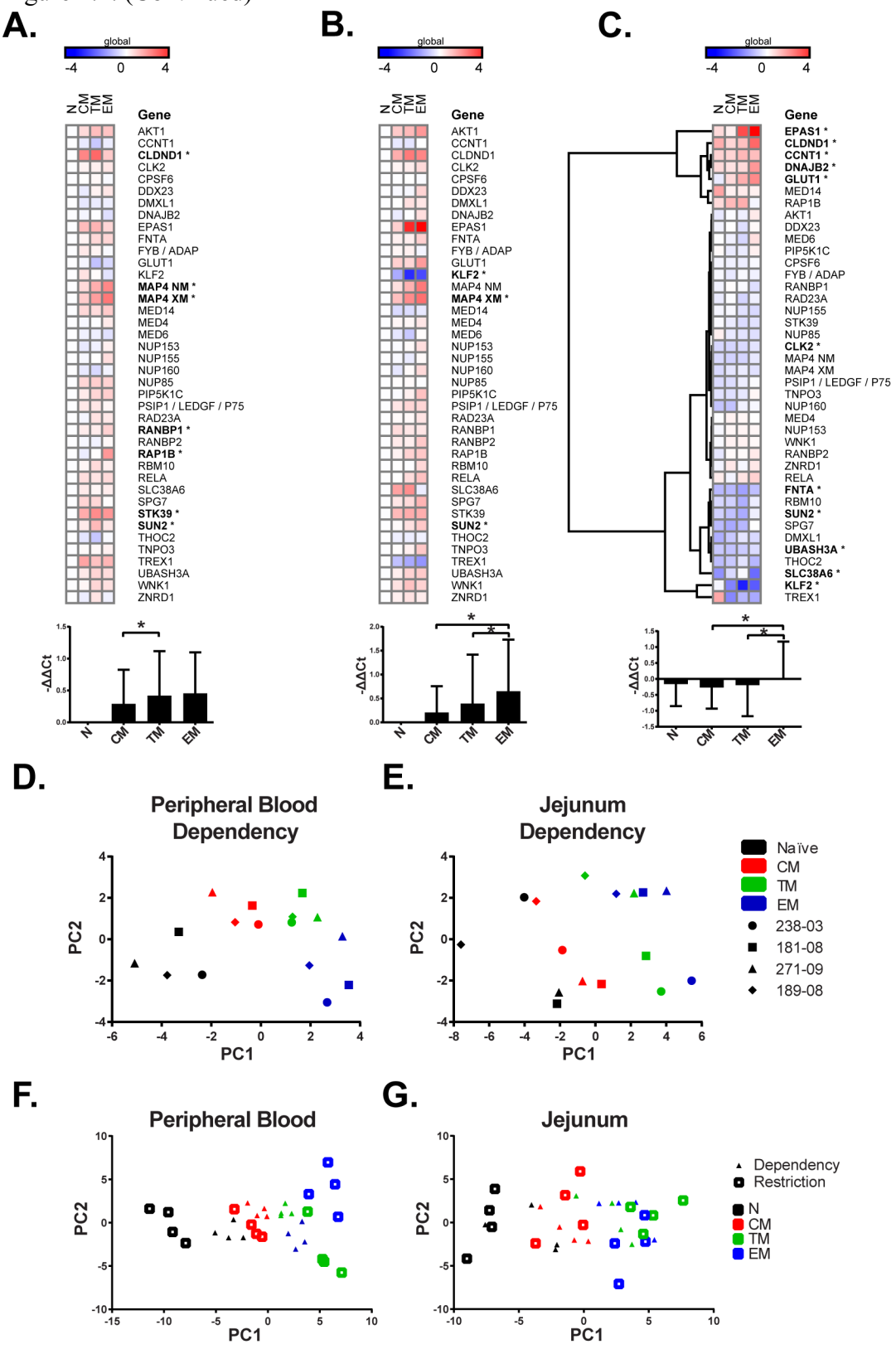
To determine if there were differences in dependency factor expression between peripheral blood and jejunum cells, expression in jejunum relative to peripheral blood was analyzed using the -ΔΔCt method, normalizing expression to the most stable endogenous control, *POLR2A*, and to the respective peripheral blood subset **(Figure 2.2.C)**. Using a two-way ANOVA, the 27.5% of genes that significantly differed in their magnitude of expression between peripheral blood and jejunum were identified and are noted in Figure 2.2.C. Additionally, *KLF2*

was the only gene identified as exhibiting a significantly different pattern of expression among the memory subsets in the two tissues.

Figure 2.2. Expression of dependency factors generally increases in memory differentiation.

Heatmap of -ΔΔCt expression values depicting memory subset expression relative to naïve for peripheral blood **(A)** and jejunum **(B)**. Means of four animals, * = $p<0.01$, repeated measures ANOVA with Benjamini Hochberg correction. Bar graphs under each heatmap depicts mean –ΔΔCt expression for each subset. * = $p<0.05$, paired T-test. Heatmap depicting expression in jejunum relative to peripheral blood and the most stable endogenous control, *POLR2A* **(C)**. Unsupervised hierarchical clustering using uncentered Pearson correlation with complete linkage groups genes with similar expression patterns. Mean of four animals, * = $p<0.05$, two-way ANOVA testing tissue differences with Benjamini Hochberg correction. Principal component analysis for peripheral blood **(D)** and jejunum **(E)** based on analysis of 40 dependency factor genes. Overlay depicting principal component analysis of restriction and dependency factors on the same scale for peripheral blood **(F)** and jejunum **(G)**.

Figure 2.2. (Continued)



Stimulation alters restriction factor expression

To test if cellular activation modulated expression of restriction factors, bulk CD4+ T cells were isolated from the peripheral blood of four rhesus macaques and stimulated with either αCD3/αCD28 coupled beads or 1000 units/ml of recombinant type I human interferon-αA/D. CD3/CD28 activation significantly ($p<0.05$, t-test) modulated 36% of the genes in at least one time point as compared to mock stimulated cells **(Figure 2.3.A)**. Stimulation induced both increases and decreases in expression of multiple restriction factors over the course of the experiment, showing that restriction factor expression can be dynamically modulated by CD3/CD28 stimulation. This pattern of modulation of expression of restriction factors contrasts with that previously observed in response to PHA stimulation, which generally resulted in increased restriction factor expression in human CD4 T cells [320].

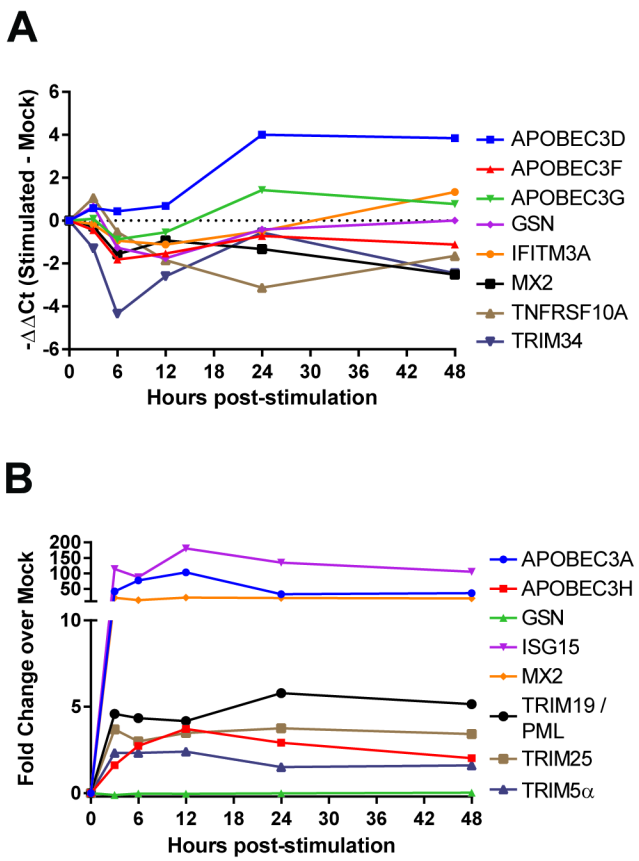


Figure 2.3. Expression of restriction factors is dynamically modulated by CD3/CD28 and type I interferon stimulation.
Time course of the change in expression of selected restriction factor genes in CD4+ T cells from peripheral blood stimulated with antiCD3/antiCD28 beads over mock as relative to the time of stimulation **(A)**. Time course of the change in expression of restriction factor genes after stimulation with 1000 units/ml of type I interferon, relative to the time of stimulation **(B)**. Displayed here is data from the assay specific for the α isoform of *TRIM5*. Mean of four animals.

In vitro interferon stimulation increased expression of 32% of the genes compared to mock-stimulated cells **(Figure 2.3.B)**. All of the restriction factors exhibiting increased expression have been previously found to be interferon-stimulated genes in microarray studies [321-323]. The time course exhibited a rapid modulation of responding genes as they reached peak or near-peak fold changes in the earliest time point post stimulation, three hours, and generally maintained these levels for 48 hours. Similar to memory differentiation, both *TRIM5* assays were

modulated in concert by interferon, indicating similar regulation of the different transcripts. Overall, both forms of stimulation showed that restriction factor expression can be dynamically modulated by external signals. However, neither form of stimulation reproduced a pattern similar to memory differentiation.

Differences between jejunum and peripheral blood restriction factor expression

Despite the similarities in the overall pattern of restriction factor expression between peripheral blood and jejunum, several differences were apparent. To determine the extent of variation in restriction factor expression between peripheral blood and jejunum CD4+ T cells, expression of restriction factors in jejunum relative to peripheral blood was analyzed using the ΔΔCt method [318], normalizing expression to the most stable endogenous control, *B2M*, and to the respective peripheral blood subset **(Figure 2.4.A)**. Several genes showed more than four-fold greater differences: *APOBEC3D* in naïve, *PRMT6* and *TNFRSF10D* in effector memory, and *ZC3H12A* in all three memory CD4+ T cell subsets. Interestingly, eight genes exhibited at least four-fold mean decreased expression in the CCR5+ transitional memory cells from jejunum as compared to the same subset of CD4+ T cells from peripheral blood. The number of restriction factors with decreased expression in jejunal CCR5+ transitional memory CD4+ T cells was much larger than the number of genes exhibiting lower expression in central memory (one) or effector memory (two). Using a two-way ANOVA, the genes that significantly differed in their magnitude of expression between peripheral blood and jejunum were identified and are noted in Figure 2.4.A. Additionally, genes for which the pattern of expression among the memory subsets differed significantly in the two tissues were identified. There were nine genes with a p-value less than 0.02: *APOBEC3*s *C*, *D*, *G*, and *H*, *APOL6*, *CDKN1A*, *PARP1*, *SAMHD1*, and *TRIM5*.

As the overall level of viral restriction mediated by these genes is likely to be cumulative, the overall mean expression of restriction factors from jejunum as relative to peripheral blood was calculated **(Figure 2.4.B)**. This value provides a measure of the extent of overall restriction factor expression in each subset for which the only difference between cells was anatomic location. Transitional memory cells from the jejunum had a mean -ΔΔCt value of -0.819, which represents 0.567 fold lower expression. Using a paired T test, this decrease was statistically significantly different from the naïve (P=0.028) and central memory (P=0.015) subsets, though not for effector CD4+ T cells (P=0.115).

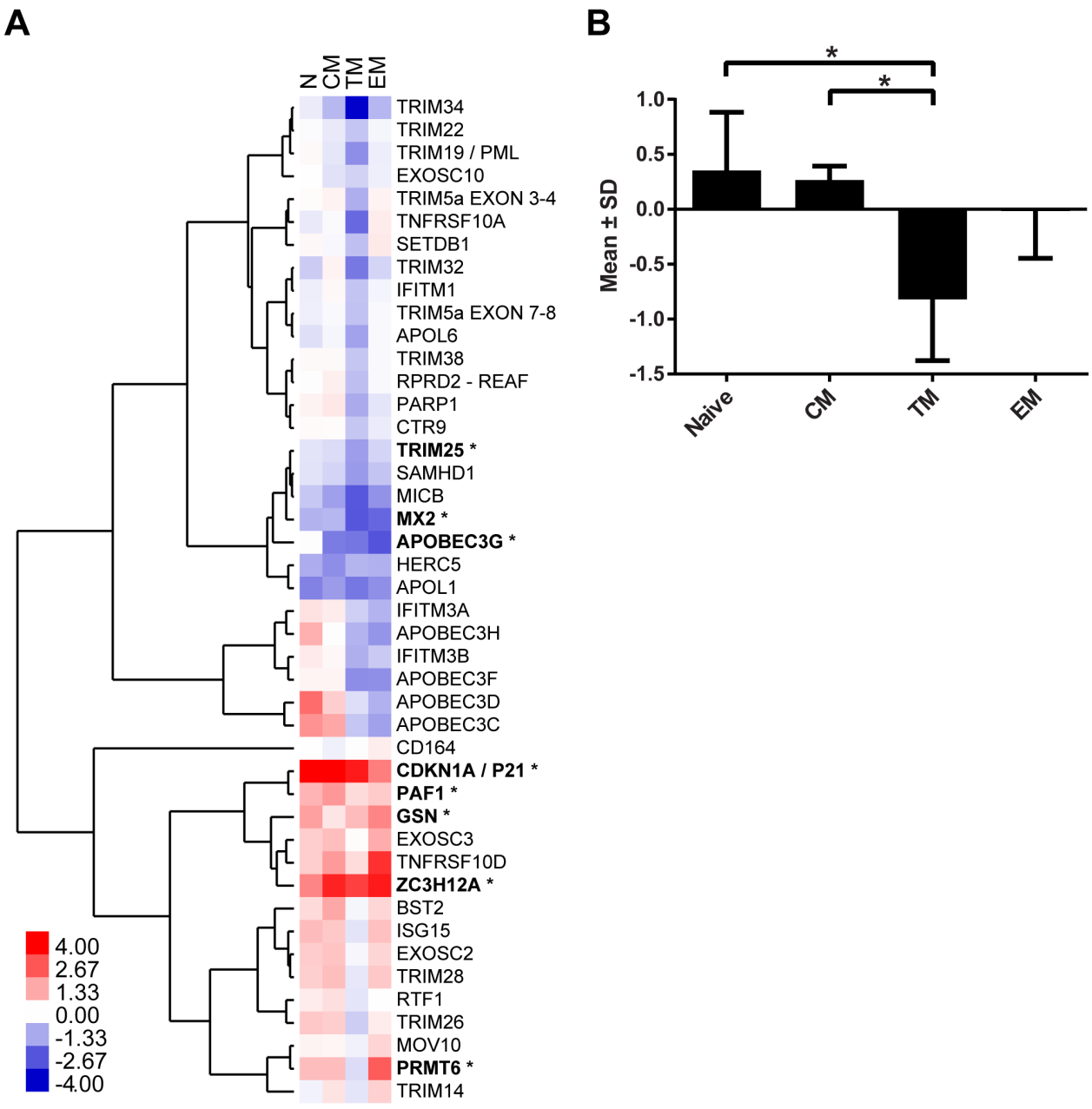


Figure 2.4. Differences in restriction factor expression in peripheral blood and jejunum CD4+ T cells.

Heatmap depicting expression in jejunum relative to peripheral blood and the most stable endogenous control **(A)**. Unsupervised hierarchical clustering using uncentered Pearson correlation with complete linkage groups genes with similar expression patterns. Mean of four animals, * = $p<0.05$, two-way ANOVA testing tissue differences with Benjamini Hochberg correction. Mean expression + SD ($-\Delta\Delta Ct$) of all restriction factors in jejunum relative to their respective peripheral blood subset **(B)**. * = $p< 0.05$, paired T test.

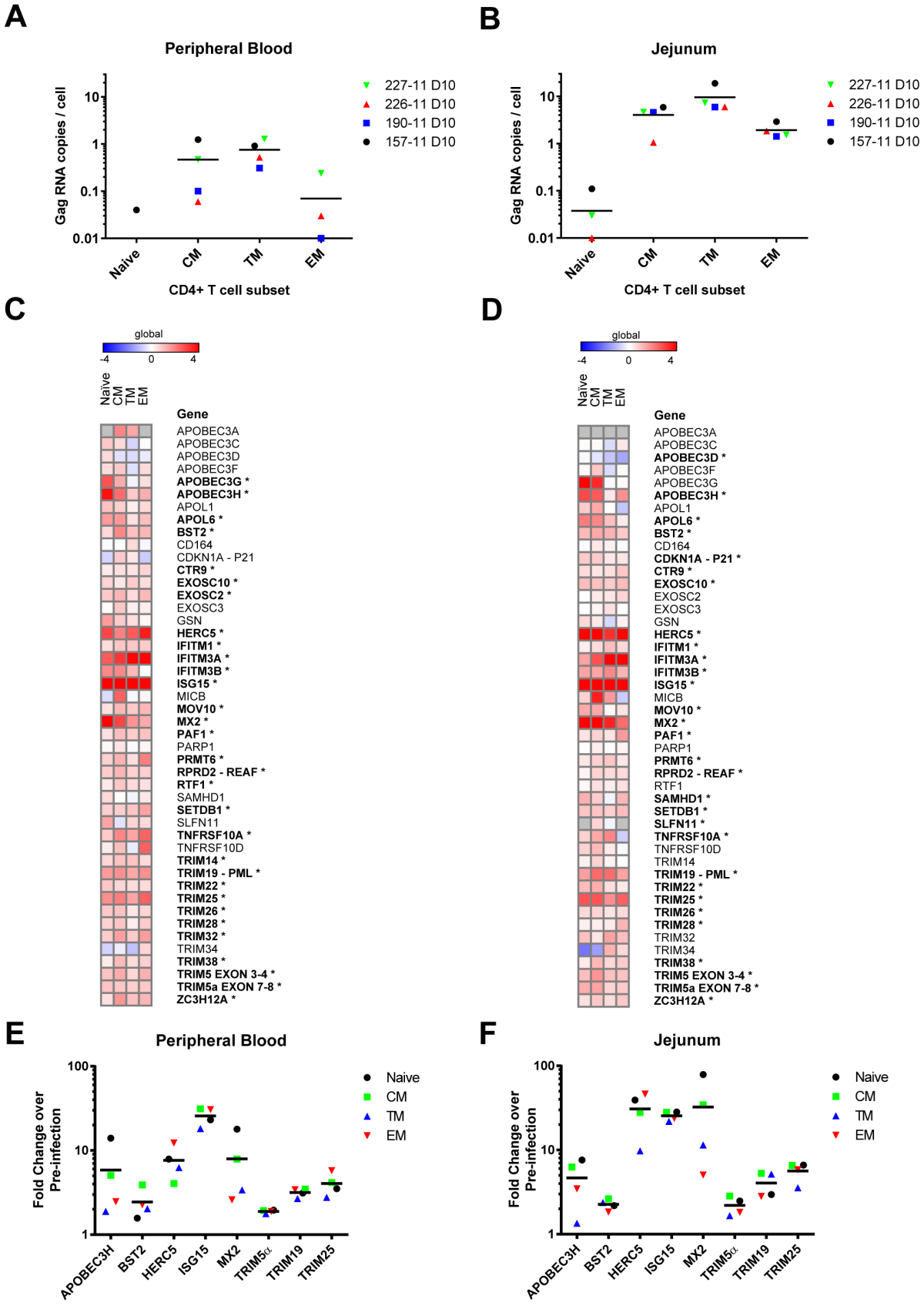
Acute SIV infection increases expression of restriction factors in CD4+ T cells

Given the effects of IFN stimulation and T cell activation that we observed on the expression of restriction factors in CD4+ T cells *in vitro*, we reasoned that the dynamic effects of acute SIV infection, which include induction of a robust type I interferon response [206,290], would likely have significant effects on the expression of restriction factors *in vivo*. It is challenging to study the acute phase of immunodeficiency virus infection in humans, especially in the primary target tissue of the GALT. To investigate potential changes in expression of restriction factors *in vivo* during acute SIV infection, four rhesus macaques were infected intravenously with $SIV_{MAC}239$. Ten days post-infection, animals were sacrificed and peripheral blood and jejunum lymphocytes were isolated. Memory CD4+ T cell subsets were sorted as before and levels of SIV *gag* RNA were quantified. In both peripheral blood and jejunum, the transitional memory CCR5+ CD4+ T cells were the most highly infected memory subset, as expected **(Figure 2.5.A and 2.5.B)**. CD4+ T cells from jejunum in each subset were much more highly infected than those from peripheral blood ($p=0.0058$, paired T test). The reduced expression of restriction factors specifically in jejunum transitional memory CD4+ T cells as compared to peripheral blood corresponded with increased viral infection of these cells.

Figure 2.5. Increased SIV infection in jejunum transitional memory CD4+ T cells and acute infection changes in restriction factor expression.

Gag RNA copies per cell equivalent were quantified in sorted CD4+ T cell subsets from peripheral blood **(A)** and jejunum **(B)**. Heatmap of -ΔΔCt expression values depicting memory subset expression in post-infection cells relative to uninfected for peripheral blood **(C)** and jejunum **(D)**. Means of 4 animals. * = $p < 0.01$, two-way ANOVA with with Benjamini Hochberg correction testing differences due to infection. Mean fold change of selected restriction factors from CD4+ T cell subsets post-infection relative to the same subset pre-infection for peripheral blood **(E)** and jejunum **(F)**.

Figure 2.5. (Continued)



Strikingly, expression of restriction factors increased in every CD4+ T cell memory subset from both peripheral blood (2.08 mean fold change increase) and jejunum (2.14 mean fold change) ten days post-infection **(Figure 2.5.C and 2.5.D)**. On average, 43 of 46 genes increased expression in peripheral blood and 43 of 45 in jejunum. In particular, peripheral blood cells exhibited a mean increase of over 4-fold in *APOBEC3H*, *HERC5*, *IFITM3A*, *ISG15*, and *MX2* while jejunum cells increased expression of *APOBEC3G*, *HERC5*, *IFITIM3A*, *ISG15*, *MX2*, and *TRIM25* more than 4-fold on average. These genes have been previously reported to be type I interferon-stimulated genes (ISGs) [322,323], and acute SIV infection has been well-documented to induce a strong type I interferon response [206]. Using a two-way ANOVA with Benjamini Hochberg correction, approximately 2/3 of restriction factors were significantly modulated ($p < 0.01$) by infection in both peripheral blood and jejunum, including the known ISGs and nearly all of the *TRIM* family members. To verify that global gene expression increase did not account for the restriction factor upregulation, a panel of 18 endogenous control or lineage genes such as *CD3*, *CD4*, *CD28*, etc., was confirmed to remain quite stable as compared to pre-infection (mean $-\Delta\Delta Ct$ value of -0.04, or a 0.97 fold change, data not shown). The pattern of restriction factor upregulation was quite similar in peripheral blood compared to jejunum and naïve compared to memory cells despite large differences in the level of infection between the two groups, implicating systemic interferon or cytokine responses as drivers of upregulation.

2.E. Discussion

Here we describe precise expression measurements for 45 confirmed or putative restriction factors and 40 dependency factors in subsets of primary CD4+ T cells from peripheral blood and the jejunum, a critical site for SIV/HIV pathogenesis, under a variety of different

conditions. Memory differentiation induced a high dynamic range of expression of restriction factors. For example, we observed over 22-fold average upregulation of *APOBEC* family genes in jejunum relative to their expression in naïve CD4+ T cells. Memory differentiation increased dependency factor expression on average, which could contribute to the increased susceptibility of these cells to infection. Paired comparison of jejunum with peripheral blood CD4+ T cells demonstrated unappreciated differences in restriction factor expression that could underlie differences in their susceptibility to SIV infection. Specifically, CCR5+ CD4+ T cells from jejunum showed reduced expression of restriction factors relative to the CCR5+ CD4+ T cells from peripheral blood.

In general, memory differentiation had similar effects on expression of restriction factors in both tissue compartments, increasing average restriction factor expression. Activation of bulk populations of CD4+ T cells has been shown to increase restriction factor expression [320], and both CD3/CD28 and interferon stimulation dynamically modulated expression of restriction factors. However, despite the overall increased expression of restriction factors in memory CD4+ T cells, they are well documented to be more susceptible to SIV and HIV infection [283,284]. This increased susceptibility is likely to reflect factors other than restriction factors that modulate infection, such as T cell activation as well as the ability of SIV to counteract the effects of restriction factors by various mechanisms, including the effect of Vif on APOBEC3G, Nef on Tetherin, and Vpx on SAMHD1 [92,324].

The Fluidigm Biomark qRT-PCR system used in this study allows for high-throughput, highly quantitative, and precise gene expression measurements with a wide dynamic range of 6-8 orders of magnitude, significantly wider than microarray based systems [325]. In addition, the specificity of the primer/probe real-time PCR assays permitted discrimination of specific

transcript isoforms. TRIM5 is known to be expressed as multiple isoforms in humans. Only the *TRIM5α* isoform containing the C-terminal SPRY domain is capable of restricting HIV/SIV while other isoforms can act in a dominant negative fashion to inactivate the α isoform [317]. Here we found that the α isoform is present at about 50% of the abundance of total *TRIM5* in primary CD4+ T cells from both jejunum and peripheral blood. However, various forms of stimulation including memory differentiation, CD3/CD28, interferon, and SIV infection did not appreciably alter the relative abundance of the α isoform to that of total *TRIM5* despite altering the total level of expression.

This study utilized the advantages of the rhesus macaque/SIV model to understand the expression changes of restriction factors in primary cells and cells from tissues that are difficult to study in humans, especially during acute infection. In particular, mucosal tissues are the primary site of transmission and the major site for replication and CD4+ T cell depletion in HIV/SIV infection [163,165,180,289]. Infection of CD4+ T cells in peripheral blood is relatively infrequent, 0.1 to 1%, while up to 60% of mucosal CD4+ T cells can be infected during acute infection [165,283]. Here we observed increased levels of infection in each CD4+ T cell subset in jejunum as compared to peripheral blood, verifying mucosal CD4+ T cells as particularly susceptible targets. Though precise mechanisms remain unclear, multiple hypotheses have been proposed to explain this increased susceptibility to SIV/HIV infection, including a greater proportion of CCR5-expressing cells, fewer naïve cells, more activated cells with higher levels of transcription, the closer proximity of target cells facilitating cell-cell spread, and additional factors such as increased α4β7 expression facilitating the binding of virions [13]. Here we show that in the CCR5-expressing transitional memory subset with the same phenotype from peripheral blood and jejunum, cells from jejunum expressed relatively lower levels of restriction

factor genes and that memory cells expressed higher levels of dependency factor genes, both potentially contributing to their increased levels of infection.

Previous studies have measured the expression of a subset of the restriction factors included in this study under certain conditions. For example, PBMCs from HIV-naïve subjects stimulated with PHA induced upregulation of genes such as *ISG15*, *APOBEC3G*, and *TRIM5* [320]. Interferon-α treatment of HIV/HCV co-infected patients resulted in the upregulation of a subset of restriction factor genes like *ISG15* and *TRIM19 (PML)* in agreement with our results [203,326]. However, these studies measured restriction factor expression in either total PBMCs or total CD4+ T cells. As shown in this study, restriction factor expression can vary significantly among different memory subsets. Therefore, analysis of expression of specific restriction factors in specific populations of target cells may be more informative than analysis in bulk populations of CD4+ T cells.

We demonstrated dynamic modulation of restriction factor expression in CD4+ T cells due to differentiation, activation, interferon stimulation, and SIV infection. All forms of stimulation induced different patterns of expression of restriction factors. For example, memory differentiation increased expression of all *APOBEC* family members, while anti-CD3/CD28, interferon, and SIV infection increased expression of only selected *APOBEC* genes. Differential effects of stimulation suggest different mechanisms of regulation. Acute SIV infection is known to induce robust upregulation of interferon expression [206,290]. Acute SIV infection increased expression of nearly every gene while *in vitro* interferon stimulation induced expression of only 32% of the measured restriction factors, suggesting type I interferon alone is insufficient to account for the full range of expression modulation during infection. Previous studies have found a wide range of cytokine/chemokine production during acute SIV/HIV infection [199,327], and

future studies will be required to determine the specific contributions of these signals to restriction factor regulation.

Recent research shows that innate immune responses may be suppressed in the first days post-infection thus allowing viral replication to progress [328]. In this study, nearly all of the restriction factors increased in expression in both peripheral blood and jejunum CD4+ T cells at 10 days post-infection, with interferon-stimulated genes exhibiting the greatest extent of upregulation. While these restriction factors may play a role in inhibiting SIV replication, this dramatic upregulation does not occur until after systemic infection has been established. The kinetics of restriction factor expression changes during the earliest phases of acute pathogenic infection are still unknown, including if cells at the earliest sites of viral replication display gene expression changes before peripheral cells. Additionally, it is still unknown if differences in restriction factor expression may underlie differences in susceptibility to infection at the organismal level.

Taken together, these data highlight the dynamic regulation of expression of restriction and dependency factors, which can be modulated by multiple factors including cell differentiation, anatomic compartment, T cell activation, and acute SIV infection. Although other factors are clearly relevant, modulation of expression of restriction and dependency factors is likely to play a significant role in the wide range of differential susceptibility of distinct CD4+ T cell subsets to SIV infection. Future studies will be necessary to better to define the roles of individual restriction or dependency factors to the susceptibility of primary CD4+ T cells to SIV infection.

Chapter 3: Identification and characterization of a highly SIV-susceptible population of Tfh-like cells in rhesus macaque jejunum by single cell gene expression analysis

Acknowledgements

We thank Francois Villinger (Emory University) for providing the $SIV_{MAC}239$ virus stock, Arnaud Colantonio for study coordination, Jackie Gillis, Michelle Connole, Barbara Cervasi, and Kiran Gill for flow cytometry support, and Jeffrey Lifson (NCI-Frederick) for providing advice on the SIV PCR primers. We thank Thomas Vanderford and Benton Lawson for plasma viral load measurements. We also thank Amber Hoggatt and other members of the New England Primate Research Center (NEPRC) Division of Veterinary Resources for their expert animal care, and Elizabeth Curran and other members of the NEPRC Pathology Division for their assistance.

3.A. Abstract

Despite extensive study, our understanding of the molecular events that determine the susceptibility of CD4+ T cells to SIV and HIV infection and the cellular events that follow lentiviral infection have been limited by our ability to track events that occur in single cells and analyze gene expression, including viral gene expression, on a single cell basis. New techniques that permit high-throughput analysis of gene expression in single cells can be used to deconvolute cell types in apparently homogenous populations of bulk cells and offer the potential to provide new insights into AIDS pathogenesis. Here we utilized acutely simian immunodeficiency virus (SIV)-infected rhesus macaques to isolate individual infected and uninfected CD4+ T cells from the intestinal mucosa, the primary site of viral replication in acute infection. Using high-throughput microfluidic quantitative real-time PCR of single cells, we measured expression of five viral transcripts used to define SIV-infected cells along with 91 cellular genes chosen for potential relevance in the viral replication cycle.

Single cell analysis of over 300 single jejunal CD4+ T cells obtained 10 days after intravenous SIV infection revealed that approximately 20% of these cells were SIV-infected. Comparison of gene expression profiles using multiple statistical methods identified PD-1 and CXCR5 as being the most significantly differentially expressed genes between infected and uninfected cells, and the coexpression of both genes was observed more frequently in infected cells.

The coexpression of PD-1 and CXCR5 on CD4+ T cells defines T follicular helper (Tfh) cells. However, the existence of a Tfh population in the jejunum has not previously been described. Using flow cytometry, we found a discreet population of PD-1+ CXCR5+ memory CD4+ T cells in the jejunum of uninfected animals with a frequency intermediate to that found in peripheral blood and lymph nodes. This jejunal PD-1+ CXCR5+ cell population expressed surface markers ICOS, CD69, CTLA4, CCR5, CD200, and the canonical Tfh transcription factor BCL6 protein at similar levels to classical Tfh populations from lymph node or spleen. Additionally, the expression profile of over 70 Tfh-associated genes in the PD-1+ CXCR5+ cells from jejunum was remarkably similar to classical Tfh cells, including the canonical Tfh cytokine IL-21. This novel population exhibits many phenotypic and transcriptional characteristics of Tfh cells, yet is found in a location that has not been generally considered to be an inductive site. While circulating Tfh cells have been described in peripheral blood, the jejunal population described in this study was found at a much higher proportion of memory CD4+ T cells and expressed BCL6 protein as well as the activation markers CD69 and ICOS, differentiating it from circulating Tfh cells.

PD-1+ CXCR5+ cells from jejunum were found to be highly SIV-infected, with an average of over three SIV DNA copies/cell at the peak of acute infection. This level of infection

higher than that of bulk memory CD4+ T cells was observed despite low levels of cell surface expression of the SIV coreceptor CCR5. This study is the first single cell gene expression analysis of primate lentivirus-infected cells, and it identified a novel and highly susceptible target cell population *in vivo* during acute infection.

3.B. Introduction

Human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV) infect CD4+ cells through their use of CD4 as a primary receptor and a chemokine receptor (the two most common are CCR5 and CXCR4) as a coreceptor for entry into cells [29,30]. Cells that express these receptors and are therefore potentially susceptible to infection include CD4+ T cells, CD4+ monocytes [329,330], macrophages [331], thymocytes [332], some subsets of dendritic cells [333], some subsets of hematopoietic progenitor cells [334,335], and CNS macrophages including microglial cells [336,337] . However, expression of the proper receptors alone is not sufficient to explain a cell's susceptibility to infection by immunodeficiency viruses. Many additional parameters have been suggested to influence susceptibility, though the relative contributions of these factors remain undefined. The level of cell surface receptor or coreceptor [338] and the level of autocrine CCR5-binding chemokine production that influences receptor availability [93] can impact susceptibility, as can the presence of α4β7 on the cell surface, which HIV and SIV are capable of binding [339,340]. Cell activation, although not well-defined molecularly, also influences whether a cell can be infected, potentially through multiple mechanisms. For example, to achieve efficient infection *in vitro* T cell-activating agents like ConA are generally used. Additionally, spinoculation, thought to work by physically concentrating virus, has recently been reported to function by increasing cell activation by actin- and cofilin-mediated responses to centrifugal stress [341]. Similarly, the

memory status of a cell (if a cell is antigen-experienced) influences its susceptibility [283,284]. The expression of restriction factors can influence susceptibility, for example the expression of SAMHD1 in non-cycling cells renders them resistant to HIV infection by depleting nucleotide pools necessary for reverse transcription [104]. Finally, other markers for susceptibility without a known mechanistic link to viral infection such as CCR6 expression (a marker for TH17 cells) have been described [342]. In total, although many single parameters contributing to HIV and SIV infection have been identified, often using *in vitro* studies, a complete molecular description of the characteristics of the subsets of CD4+ T cells that are susceptible to HIV/SIV infection *in vivo* is still lacking. To that end, this study focused on primary cells and simultaneously analyzed multiple factors in individual cells.

Accumulating evidence emphasizes that the mucosal immune system, and specifically the intestinal immune system, is central to the HIV/SIV infection course. The majority of natural transmissions of HIV are across a mucosal barrier and the virus primarily replicates in and depletes gut CD4+ T cells [163,165,166,180]. Despite the clear importance of mucosal immune tissue in the initiation and progression of infection, cells from mucosal tissues are relatively understudied as compared to cells from peripheral blood due to the difficulty in obtaining tissue samples [289]. This emphasizes the utility of non-human primates in studying immunodeficiency virus infections since the time of infection is precisely known and sufficient tissue samples can be obtained. Additionally, there are many similarities in the viruses, hosts, and disease pathogenesis between HIV in infection in humans and SIV in rhesus macaques [164].

Widely used methods to measure gene expression such as microarrays and quantitative real-time PCR (qRT-PCR) generally analyze relatively large numbers of cells (10^6 or more). As a result, analysis of gene expression in these studies reflects an average of all the cells in the

sampled population. Gene expression on a single cell level is variable such that even housekeeping genes in genetically identical cells have a broad range of gene expression values [278]. This variability can be due to the intrinsic stochastic nature of biochemical processes regulated by a small number of molecules and extrinsic sources such as the cell cycle and the unique signals from the local microenvironment [277,279]. Therefore, measurements that average expression over multiple cells can obscure differences among subpopulations. Single cell measurements can deconvolute cell types in putatively homogenous populations to discover the identity and gene expression profiles of subpopulations, even if they are rare [278].

Multiple studies have attempted to characterize differences in gene expression between HIV-infected cells and uninfected cells [343-347]. Potential confounding factors *in vitro* include use of cell lines; infection with super-physiological levels of virus resulting in infection of cells that may not be infected *in vivo*; cell stimulation to increase infection rates which alters gene expression; and using mock-infected cells rather than cells that do not become infected as a comparison to virus-exposed cells. In *in vivo* studies, gene expression measurements are averages of many cells. As even during the peak of infection in highly susceptible mucosal tissue only about 50% of memory CD4+ T cells are infected [165], bulk gene expression measurements are averages of both infected and uninfected cells. No study to date has measured gene expression on a single cell basis in HIV- or SIV-infected CD4+ T cells as compared to individual uninfected cells.

T follicular helper cells (Tfh) are a subset of CD4+ helper T cells specialized for providing help for B cells [152]. They are essential for germinal center formation, affinity maturation, class-switch recombination, and allow the formation of plasma cells and long-lived memory B cells. A number of recent publications have resulted in a rapid expansion of our

understanding of the importance of Tfh cells in HIV/SIV infection. In chronic untreated HIV-infected patients Tfh cells were found to be the major T cell compartment for HIV infection, replication, and production [269], and an important site of viral persistence in chronic infection [271]. Tfh cells are generally located in germinal centers in secondary lymphoid tissues and can be found in low frequency in the periphery, though these cells express lower levels of CXCR5, BCL-6, ICOS, and PD-1 [348,349]. There is only limited evidence for Tfh cells in non-lymphoid mucosal tissue. One of the only studies looking for PD-1^{high} cells in jejunum failed to find any during chronic SIV infection, suggesting that either they are depleted or that they are not present due to limited GALT structures in jejunum [350]. Limited recent evidence suggests Tfh or Tfh-like cells can exist in mucosa sites. Mycobacterium tuberculosis-specific PD-1^{high} cells residing in the lung parenchyma required intrinsic Bcl6 and ICOS expression for their generation, and a fraction of these cells co-expressed CXCR5 [351]. Additionally, in a humanized-DRAG mouse model human Tfh cells were found in gut lamina propria at about 1/3 the frequency of that in Peyer's patches or mesenteric lymph nodes [352].

In this study, we utilized high-throughput single cell RT-PCR to analyze gene expression from single jejunum CD4+ T cells from acutely SIV-infected rhesus macaques. Comparison of gene expression between SIV-infected and uninfected T cells revealed PD-1 and CXCR5 as the most significantly differentially expressed genes, and both genes were enriched in SIV-infected cells. Together PD-1 and CXCR5 can define Tfh cells; however, Tfh cells have not been described in jejunum tissue lacking organized secondary lymphoid tissue. A PD-1+ CXCR5+ population of memory CD4+ T cells was identified by flow cytometry in the jejunum of uninfected animals, and these cells were found to express many phenotypic markers associated with Tfh cells at similar levels to Tfh cells isolated from spleen or lymph node. Additionally,

they expressed the canonical Tfh transcription factor BCL6 protein and IL-21 mRNA. To rigorously compare the gene expression profile of this novel population to classical Tfh cells from secondary lymphoid tissues, we examined the expression of 77 Tfh-associated genes in PD-1+ CXCR5+ cells from jejunum and lymph nodes. PD-1+ CXCR5+ cells from jejunum exhibits a characteristic Tfh signature of gene expression, clustering with Tfh cells from lymph nodes by hierarchical clustering and principal component analysis. The PD-1+ CXCR5+ cells from jejunum were found to contain high levels of SIV DNA and RNA during acute infection, showing these cells are particularly susceptible to infection and suggesting that they are depleted rapidly.

3.C. Materials and Methods

Animals

The eight Indian-origin rhesus macaque monkeys (Macaca mulatta) described in this study were housed at the New England Primate Research Center (NEPRC) in accordance with the regulations of the American Association of Accreditation of Laboratory Animal Care and the standards of the Association for Assessment and Accreditation of Laboratory Animal Care International. All protocols and procedures were approved by the relevant Institutional Animal Care and Use Committee. Care met the guidance of the Animal Welfare Regulations, OLAW reporting, and the standards set forth in The Guide for the Care and Use of Laboratory Animals [291]. Euthanasia took place at seven or ten days post-infection using protocols consistent with the American Veterinary Medical Association (AVMA) guidelines.

Lymphocyte isolation

Jejunum tissue was isolated at time of euthanasia, washed with PBS, separated into small pieces with a scalpel, incubated with 5 mM EDTA for 30 min, washed and incubated with 15 U/ml type II collagenase (Sigma-Aldrich) followed by mechanical disruption with an 18 gauge-feeding needle and filtration through 70 μm cell strainers (BD Biosciences). Lymphocytes were then enriched by bilayer (35%/60%) isotonic Percoll density gradient centrifugation (1000 × g, 20 min), and the interface containing the lymphocytes was collected [292].

Spleen lymphocytes were isolated by briefly disrupting the tissue with a scalpel, filtering through a 70 um cell strainer, lysing red blood cells with ACK buffer, and washing.

Lymph node cells were isolated by disrupting the capsule, filtering through a 70 μm cell strainer, and washing.

Peripheral blood mononuclear cells (PBMC) were separated by density gradient centrifugation (Lymphocyte Separation Medium; MP Biomedicals Inc., Solon, OH) and red blood cell lysis with ACK buffer. Cells were cryopreserved in FBS media containing 10% DMSO.

Antibodies and cell sorting

To sort memory CD4+ T cells for single cell gene expression analysis, PBMC were stained with CD3 (SP34)-Pacific Blue, CD4 (L200)-FITC, CD8 (RPA-T8)-Alexa 700, CD28 (28.2)-ECD (Beckman Coulter), CD95 (DX2)-APC, CCR5 (3A9)-PE, and CCR7 (2D12)-PE-Cy7. Antibodies were obtained from BD Pharmingen unless otherwise specified. PBMCs were initially labeled with LIVE/DEAD viability stain (Life Technologies) and washed; incubated with the CCR7 antibody for 15 min at 37°C, and then incubated with all other antibodies for 20

min at room temperature and washed prior to sorting. Cell sorting was performed using a FACS Aria II cell sorter (BD Biosciences). Sorted cell populations were > 99% pure.

A phenotyping panel for Tfh-like cells included Live/Dead stain, CD3(SP34.2)-BV650, CD4(L200)-APC-Cy7, CD8(RPA-T8)-Alexa 700, CD28(CD28.2)-FITC, CCR5(3A9)-APC, CCR7(150503)-BV711, CD200(MRCOX104)-PerCP-Cy5.5, CTLA4(BN13)-PE-Cy5, CD95(DX2)-PE-Cy7 (eBiosciences), CXCR5(MU5UBEE)-PE (eBiosciences), PD-1(EH12.2H7)-BV421 (BioLegend), CD278/ICOS(398.4A)-PerCP-Cy5.5 (BioLegend), and CD69(TP.55.3)-ECD (Beckman Coulter). Additional antibodies used included BCL6(K112-91)-Alx-647, CD200(MRCOX104)-PerCP, and CCR7(150503)-ECD.

Intracellular BCL6 was performed by washing cells with PBS and staining with Live/Dead dye. After cell surface staining, cells were treated with 1mL Transcription Factor Fix/Perm(1X) solution (Tonbo Bioscience) for 30 min at 4°C and then washed 2 times with Flow Cytometry Perm Wash buffer (1X) (Tonbo). Cells were incubated with 5uL antibody recognizing nuclear antigen BCL6(K112-91)-Alx-647 at 40°C for 30 mins.

Flow cytometry was performed on a LSRII cytometer (Becton Dickinson) and analyzed with FlowJo V10 software.

Gene expression analysis

Real-time PCR assays specific for the indicated rhesus macaque genes were purchased from Applied Biosystems (ABI) **(Table 3.1)**. When not commercially available, macaque-specific custom assays with FAM-MGB probes were designed using Primer Express 3.0 (ABI). Sequences for custom assays were: *APOL6* F-AGGCAGAGGAAGAAAGTGAAGCT, R-TCGTCTTCACACAGAGGAACATCT, P-TTGGTTTGGAAAGGGATGAG; *CXCR4*

F- ACCGCGTCTGGAGAACGA, R- TCCCCTGAGCCCATTTCC, P-
ATGGAGGGGATCAGTATATA; *CXCR5* F- CCGCTAATGCTGGAAATGGA,
R- TGTCGTTATAGTTGTCGAATTTGTCA, P- TGGAGGACCTGTTCTT; *IFITM3A*
F-GGCCAGCCTCCCAACTATG, R-GGGCGCCCCCATCAT, P-AAGAGCACGATGTGGC;
IFITM3B F-AAACCGTCTTCCCTCCTGTCA, R-GCTACCTCATGCTCTTCCTTGAG,
P-CCCCCCAGCTATGAG; *PIP5K1C* F- TGCACATTGGCATCATTGATATC,
R- CCAGGTGTGCTCCAGCTTCT, P- CAATCCTACAGGTTCATC; *RORC*
F- TGAGAAGGACAGGGAGCCAA, R- CCACAGATTTTGCAAGGGATCA,
P- TCATGAGAACACAAATTGA; *RGS1* F- CAGATGCTGCTAAACAAATCAATATTG,
R- AGGGTTGGTGCTTTAATCTTCTTG, P- TTCCGCACTCGAGAAT; *TRIM14*
F-CAGCCAGGAGCCTGATCCT, R-CCGCTGCATCTCCTGCTT,
P-AGAGGCTTCAGGCATACA; *SPG7* F- AGAAGAATAAGGAGAAGGATAAGTCGAA;
TRIM32 F-GGCCTCAATCTGGAGAATCG, R-AACCAATGGAAAAGCCACCTT,
P-CAGAATGAGCACCACCTG; *UBASH3A* F- GAAGCCAGCAGCAGATGCA,
R- CCACCAGCACGCTCTTCCT, P- CCTTGCAGGCCACCG.

All assays were confirmed to exhibit linear amplification in an eight point, three-fold dilution series of rhesus PBMC cDNA.

SIV assay design and optimization

In order to distinguish infected from uninfected cells, real-time PCR assays were designed specific for SIVmac239 sequences. Two classes of assay were used: one that detects sequences present in the viral genome, including Gag, LTR, and R region sequences, and one that spans splice junctions such that the recognition sequence is present only in spliced products

of viral transcription. These included a singly spliced Env transcript and the multiply spliced Tat-Rev transcript. As there are significantly more viral particles that do not successfully complete a replication cycle than those that do [353], detecting spliced transcripts in a cell identifies a cell that has been productively infected. Three of the SIV assays were generously provided by Jeff Lifson (NCI Frederick): Gag, LTR, and Tat-Rev.

Primer sequences for the SIV assays were: SIV gag
F-GUCUGCGUCAUCUGGUGCAUUC, R-CAAAACAGAUAGUGCAGAGACACCUAGUG,
P- CGCAGAAGAGAAAGUGAAACACACUGAGGAAG; SIV LTR
F- CTAACCGCAAGAGGCCTTCTT, R- TTGTGGAAAGTCCCTGCTGTT,
P- ACATGGCTGACAAGAAGGAAACTCGCTG; SIV R Region
F- GGAGAGGCTGGCAGATTGAG, R- ACCCAGGCTCTACCTGCTAGTG,
P- CCTGGGAGGTTCTC; SIV major splice F- TTGCAGAGGCGGATGCAT,
R- GCGGTATAGCTGAGAGAGGATTTC, P- CACTCCAGAATCGGCCA; SIV Tat-Rev
F- GCTAAGGCTAATACATCTTCTGCATC, R- CCGTCTCTTTCTTTGCCTTCTC,
P- AACAGGACCCGGCACTGCCAAC.

All SIV-specific assays were tested for sensitivity and specificity in bulk *in vitro* infected CD4+ T cells using the same RNA extraction, cDNA conversion, and qRT-PCR protocols to ensure compatibility with cellular gene qRT-PCR assays measured using the Fluidigm Biomark system.

Acute SIV infection

Eight Indian-origin rhesus macaques were infected intravenously with a 500 50% tissue culture infective dose ($TCID_{50}$) of SIVmac239 (generously provided by Francois Villinger

(Emory University)). Four animals were sacrificed at day seven post infection and four animals were sacrificed at day 10 post infection. PBMCs were isolated pre-inoculation and a CD4 phenotyping flow panel was used to analyze CD4 subset frequencies. Cells and tissue were saved from blood, spleen, axillary LN, mesenteric LN, genital LN, bone marrow, and jejunum at the time of sacrifice. The phenotyping flow panel was run on fresh PBMC, spleen, axillary LN, mesenteric LN, and jejunum cells from the day seven and 10 animals.

Single Cell qRT-PCR

To obtain a pure population of CD95+ memory CD4+ T cells from jejunum, cells were sorted in a FACS Aria II cell sorter (BD Biosciences) into cold media. Sorted cell populations were > 99% pure. Cells were pelleted, resuspended at 1000 cells/µl, and final concentration of 60% cell suspension reagent was added. Approximately 5 µl (2000 cells) were loaded into a primed 5-10 µm capture site size integrated fluidic circuit (IFC) for the C1 Single Cell Autoprep System (Fluidigm), and stained with calcein AM live/EtBr dead dye using the cell stain script. Each capture site was visualized and the number of cells captured and their viability status recorded. The system performed lysis, reverse transcription, and specific target pre-amplification for 18 cycles with 0.2x concentration primer mix. The resulting cDNA was harvested, diluted 1:3 with DNA dilution reagent, and prepared for qRT-PCR. Output was analyzed on a Fluidigm BioMark microfluidic real-time PCR system using 96.96 dynamic arrays. 4.5 µl of diluted cDNA was mixed with TaqMan Universal Master Mix II with UNG (ABI) and a 1x final concentration of 20 GE loading buffer. 7 µl of each cDNA sample was loaded into sample wells of the 96.96 dynamic array while 20x concentration RT-PCR assays with loading reagent were added to the

assay wells. Initial calculations of cycle thresholds (Ct) were performed using the Fluidigm BioMark software version 4.1.3.

Plasma viral loads

Plasma viral RNA levels were quantitated using real-time PCR, and SIV RNA copy number was determined by comparison to an external standard curve [354,355].

Cell-associated SIV DNA and RNA

DNA and RNA were extracted from sorted cell populations using an AllPrep DNA/RNA Micro Kit (Qiagen). A DNA standard with a known copy number was used to quantify Gag DNA copies, and a genomic CCR5 standard was used to quantify the copies of CCR5 in the sample. This verified the cell numbers in each sample and allowed for calculations of SIV Gag DNA copies per cell. Samples were tested on an ABI 7900 quantitative real-time PCR instrument and analyzed with ABI SDS 2.4 software.

RNA was quantified in reference to a SIV Gag standard curve of known copy number composed of *in vitro* transcribed RNA converted to cDNA. Both the standard curve and the cellular sample cDNAs were tested with a SIV Gag qRT-PCR assay and measured using the Fluidigm Biomark system.

Statistical analysis

Statistics and graphing were performed using GraphPad Prism version 6.05 for Windows (GraphPad Software, La Jolla California) or R [293]. Heatmaps were generated using GENE-E

(http://www.broadinstitute.org/cancer/software/GENE-E/index.html) and principal component analysis coordinates were obtained using MEV version 4.9 [356] and visualized in Prism.

When comparing gene expression at the single cell level, normal statistical distributional assumptions are violated, and regular Poisson or negative binomial regression models cannot fit the data effectively. We implemented a zero-inflated Poisson (ZIP) regression approach under the guidance of Wenjun Li, a biostatistician at the University of Massachusetts Medical School. This provides three p-values, one for the odds ratio of discrete expression, one for the rate ratio of continuous expression, and a joint test.

3.D. Results

Cellular gene assay selection

As qRT-PCR requires selection of target genes, a panel of about 90 rhesus genes was chosen on the basis of likely involvement in the SIV lifecycle, differential expression, and detectable expression in primary rhesus CD4+ T cells. Categories of genes selected for analysis included: genes that could define cell lineages such as *CD3*, *CCR7*, and *CCR6*; viral entry genes such as *CD4* and *CCR5*; lymphocyte activation associated genes such as *NFKB1*; restriction factor genes such as *TRIM5* and *APOBEC3G*; dependency factor genes such as nucleoporin and mediator complex genes; and genes whose expression changed upon HIV infection [346].

Many restriction factors - genes that serve as an intrinsic host defense against virus infection - have well-described activity against primate lentiviruses, including the APOBEC3 DNA deaminase family, the TRIM family, *BST-2*/Tetherin, and *SAMHD1* [92]. Screens have also been performed to identify other restriction factors. A whole genome siRNA screen identified putative restriction factors such as the PAF1 complex and exosome components [119]. A screen for

genes sharing genomic characteristics of known restriction factors identified *APOL* and *TNFRSF* family members [287].

Host dependency factors are genes that the virus uses to complete its replication cycle. A number of these are well-described, such as receptors for entry, proteins involved in nuclear import (nucleoporin genes) [55], or proteins involved in integration such as LEDGF [357]. Genes that have been shown to be involved in the HIV or SIV lifecycle were prioritized for inclusion in our panel. Dependency factors have been identified in multiple genome-wide siRNA screens that resulted in thousands of total hits [15-18]. Since it is not practical to analyze expression of thousands of genes using high-throughput single cell RT-PCR, it was necessary to identify a subset of these putative dependency factors for analysis. Since the screens used different target cells, viruses, and siRNA libraries, genes that were identified on two or more screens are less likely to represent off target results, Many of these genes have been included in the list. Additionally, a follow-up study to one of the RNAi screens [17] has verified a set of 108 genes that have been scored as hits using multiple siRNA libraries in which the genes scored positive using at least two independent libraries [21].

To rationally pick the most differentially expressed genes from the dependency factors and other gene sets, we used Gene Set Enrichment Analysis (GSEA). GSEA is a computational method used to determine if a set of genes is statistically significantly enriched in a given ranked list [358]. There are two necessary components: a large list of genes ranked by some biological criteria (such as high to low expression) and a probe set of genes. To generate the ranked list, gene expression in human naïve, central memory, and effector memory $CD4^+$ T cells from a genome-wide microarray gene expression study [359] was used and ranked by an F-test. The F-test is an ANOVA statistic to compare the means of multiple groups at the same time. Ranking all

22,944 probes in this data set by F-test result generated a list of the genes from the most differentially expressed to the least differentially expressed in these cell types. To probe the ranked list using GSEA, the dataset of 108 genes identified as putative host dependency factors was used. The GSEA algorithm walks down the ranked list and increases a weighted running statistic (the enrichment score or ES) when a gene on the ranked list is found in the probe set and decreases the ES when the ranked list gene is not on the probe set. The algorithm generates two results. First, it determines whether the probe set as a whole is statistically significantly enriched on the ranked list, and second, the algorithm reports the point at which the Enrichment Score reached its peak and started declining. The genes in the probe set that contributed to the rising peak of the ES are called the leading edge. These leading edge genes represent the most differentially expressed genes in the probe set; for a set of 108 dependency factors, this leading edge was 22 genes.

Using the same F-test ranked list of gene expression in CD4+ T cell subsets, the ranked list was probed with a list of 118 putative HIV restriction factors [119] generating a leading edge of 37 genes. Additionally, GSEA was used to narrow down a list of NF-kB responsive genes generating a leading edge of 63 genes, and a list of 498 genes that were differentially expressed in HIV infected versus uninfected CD4+ T cells [346] generating a leading edge of 92 genes. Genes from each category were selected, assays chosen, and gene expression measured in bulk CD4+ T cell memory subsets with or without stimulation. Genes with differential expression between conditions and detectable expression in 200 cell equivalents were prioritized for inclusion in the final panel **(Table 3.1)**.

Table 3.1. List of genes selected for single cell expression analysis.
Assay ID numbers are provided for the ABI TaqMan assays; sequences for the custom assays are provided in the text. DEG = Differentially Expressed Gene.

Gene Name	Category	ENTREZ ID Mm	ENTREZ ID human	HGNC Full Name	Assay ID	Ref
BCL6	Lineage	708736	604	B-cell CLL/lymphoma 6	Rh02839507_m1	
CCR6	Lineage	574335	1235	C-C motif chemokine receptor 6	Rh02788181_s1	
CCR7	Lineage	574231	1236	C-C motif chemokine receptor 7	Rh03985963_s1	
CD28	Lineage	705313	940	CD28 molecule	Rh02621766_m1	
CD3e	Lineage	699467	916	CD3e molecule	Rh01062242_m1	
CXCR5	Lineage	701792	643	C-X-C motif chemokine receptor 5	Custom	
GATA3	Lineage	713840	2625	GATA binding protein 3	Rh02830714_m1	
IL17A	Lineage	708123	3605	interleukin 17A	Rh02621750_m1	
PDCD1 / PD-1	Lineage	100135775	5133	programmed cell death 1	Rh03418231_m1	
RGS1	Lineage	712754	5996	regulator of G-protein signaling 1	Custom	
RORC	Lineage	717052	6097	RAR related orphan receptor C	Custom	
TBX21 / T-bet	Lineage	694044	30009	T-box 21	Rh02621772_m1	
CCR5	Lineage / Entry	735311	1234	C-C motif chemokine receptor 5	Rh03043152_s1	
CD4	Lineage / Entry	713807	920	CD4 molecule	Rh02621720_m1	
CXCR4	Lineage / Entry	707329	7852	C-X-C motif chemokine receptor 4	Custom	
GPR15	Lineage / Entry	698718	2838	G protein-coupled receptor 15; BOB	Rh02786977_s1	[31]
Fos	Activation	702077	2353	Fos proto-oncogene, AP-1 transcription factor subunit	Rh01119267_m1	
LGALS3	Activation	697290	3958	lectin, galactoside binding soluble 3	Rh02841801_m1	
NFKB1	Activation	710324	4790	nuclear factor kappa B subunit 1	Rh00765726_m1	
S100A4	Activation	715115	6275	S100 calcium binding protein A4	Rh02818677_m1	
TNFRSF1B	Activation	715454	7133	TNF receptor superfamily member 1B	Rh02837327_m1	
VIM	Activation	705289	7431	vimentin	Rh00958113_m1	
APOBEC3G	Restriction	574398	60489	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	Rh02788475_m1	[115,116,299]
APOL6	Restriction	693593	80830	apolipoprotein L, 6	Custom	[95,287,360]
CDKN1A	Restriction	719199	1026	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	Hs00355782_m1	[302,303]
EXOSC3	Restriction	716347	51010	exosome component 3	Rh02830694_m1	[119]
GSN	Restriction	699705	2934	gelsolin (amyloidosis, Finnish type)	Rh02794823_m1	[304]
IFITM1	Restriction	697687	8519	interferon induced transmembrane protein 1 (9-27)	Rh02809735_gH	[94]
IFITM3A	Restriction	697829	10410	[predicted] interferon induced transmembrane protein 3-like	Custom	[94]
IFITM3B	Restriction	697564	10410	[predicted] interferon induced transmembrane protein 3-like	Custom	[94]
ISG15	Restriction	700141	9636	ISG15 ubiquitin-like modifier	Rh02915441_g1	[305]
MX2	Restriction	780935	4600	myxovirus (influenza virus) resistance 2 (mouse)	Rh02801425_m1	[131,133]
PAF1	Restriction	697786	54623	Paf1, RNA polymerase II associated factor, homolog (S. cerevisiae)	Rh02876261_m1	[119]

Table 3.1. (Continued)

Gene Name	Category	ENTREZ ID Mm	ENTREZ ID human	HGNC Full Name	Assay ID	Ref
PRMT6	Restriction	694284	55170	protein arginine methyltransferase 6	Rh02817860_s1	310,311
RPRD2 - REAF	Restriction	715526	23248	regulation of nuclear pre-mRNA domain containing 2; REAF	Rh02868968_m1	119,120
SAMHD1	Restriction	709060	25939	SAM domain and HD domain 1	Rh02869977_m1	104,107
TNFRSF10A	Restriction	716826	8797	tumor necrosis factor receptor superfamily, member 10a	Rh02846752_m1	287
TRIM14	Restriction	715418	9830	tripartite motif-containing 14	Custom	95
TRIM19 – PML	Restriction	700379	5371	promyelocytic leukemia	Rh03043124_m1	312
TRIM22	Restriction	713814	10346	tripartite motif-containing 22	Rh02801450_m1	129,313,314
TRIM32	Restriction	702595	22954	tripartite motif-containing 32	Custom	315
TRIM34	Restriction	100568287	53840	TRIM6-TRIM34 readthrough transcript; tripartite motif-containing 6; tripartite motif-containing 34	Rh04256228_m1	316
TRIM38	Restriction	694861	10475	tripartite motif-containing 38	Rh02860500_m1	127
TRIM5_3_4	Restriction	574288	85363	tripartite motif containing 5; exon 3-4 junction, all isoforms	Rh02788627_m1	121,317
TRIM5A_7_8	Restriction	574288	85363	tripartite motif containing 5; exon 7-8 junction, alpha isoform only	Rh02788631_m1	121,317
ZC3H12A	Restriction	713604	80149	zinc finger CCCH-type containing 12A; MCPIP1	Rh02882632_mH	139
CLDND1	Dependency	698472	56650	claudin domain containing 1	Rh01026490_m1	
DMXL1	Dependency	698117	1657	Dmx like 1	Rh02857893_m1	
DNAJB2	Dependency	702936	3300	DnaJ heat shock protein family (Hsp40) member B2	Rh02800276_m1	
FNTA	Dependency	712681	2339	farnesyltransferase, CAAX box, alpha	Hs00954344_g1	
FYB / ADAP	Dependency	693951	2533	FYN binding protein	Rh02838822_m1	
KLF2	Dependency	719492	10365	Kruppel like factor 2	Rh02927457_mH	
MAP4 NM	Dependency	100423619	4134	microtubule associated protein 4	Rh02841978_m1	
MAP4 XM	Dependency	711097	4134	microtubule associated protein 4	Rh02841981_m1	
MED4	Dependency	704644	29079	mediator complex subunit 4	Rh02800666_m1	
MED14	Dependency	699395	9282	mediator complex subunit 14	Rh02847948_m1	
NUP153	Dependency	703888	9972	nucleoporin 153	Rh02856816_m1	
NUP155	Dependency	706019	9631	nucleoporin 155	Rh00897468_m1	
PIP5K1C	Dependency	721744	23396	phosphatidylinositol-4-phosphate 5-kinase type 1 gamma	Custom	
PSIP1 / LEDGF / p75	Dependency	664733	11168	PC4 and SFRS1 interacting protein 1	Rh02851085_m1	
RANBP1	Dependency	719448	5902	RAN binding protein 1	Rh00862133_g1	
RANBP2	Dependency	694823	5903	RAN binding protein 2; NUP358	Hs01108576_m1	
SLC2A1 / GLUT1	Dependency	698931	6513	solute carrier family 2 member 1	Rh02861051_m1	
SLC38A6	Dependency	704017	145389	solute carrier family 38 member 6	Hs01553968_m1	
SPG7	Dependency	699488	6687	SPG7, paraplegin matrix AAA peptidase subunit	Custom	

Table 3.1. (Continued)

Gene Name	Category	ENTREZ ID Mm	ENTREZ ID human	HGNC Full Name	Assay ID	Ref
STK39	Dependency	704855	27347	serine/threonine kinase 39	Rh01085344_m1	
SUN2	Dependency	704914	25777	Sad1 and UNC84 domain containing 2 [Homo sapiens	Hs00992164_m1	
UBASH3A	Dependency	722456	53347	ubiquitin associated and SH3 domain containing A	Custom	
ZNRD1	Dependency	712151	30834	zinc ribbon domain containing 1	Rh01548529_m1	
IFI27	IFN-induced	700513	3429	interferon alpha inducible protein 27	Rh02808321_mH	361
ISG20	IFN-induced	698376	3669	interferon stimulated exonuclease gene 20	Rh01088632_m1	361
MX1	IFN-induced	715186	4599	MX dynamin like GTPase 1	Rh02842279_m1	361
OAS1	IFN-induced	712265	4938	2'-5'-oligoadenylate synthetase 1	Rh00973637_m1	361
OAS3	IFN-induced	712342	4940	2'-5'-oligoadenylate synthetase 3	Rh02859894_m1	361
RSAD2	IFN-induced	708720	91543	radical S-adenosyl methionine domain containing 2; Viperin	Rh00369813_m1	361
CCL20	DEG in HIV	574182	6364	C-C motif chemokine ligand 20	Rh02788116_m1	346
DUSP4	DEG in HIV	714290	1846	dual specificity phosphatase 4	Rh01027784_m1	346
FOSB	DEG in HIV	715896	2354	FosB proto-oncogene, AP-1 transcription factor subunit	Rh02861651_m1	346
HLA-DRA / MAMU-DRA	DEG in HIV	100187587	3122	major histocompatibility complex, class II, DR alpha	Rh02621773_m1	346
IL12RB2	DEG in HIV	700677	3595	interleukin 12 receptor subunit beta 2	Rh02839111_m1	346
LDLRAP	DEG in HIV	713032	26119	low density lipoprotein receptor adaptor protein 1	Rh02870367_m1	346
PECAM1	DEG in HIV	718302	5175	platelet and endothelial cell adhesion molecule 1	Rh02828627_m1	346
PIK3IP1	DEG in HIV	716694	113791	phosphoinositide-3-kinase interacting protein 1	Rh02851758_m1	346
RALGPS2	DEG in HIV	717165	55103	Ral GEF with PH domain and SH3 binding motif 2	Hs01000187_m1	346
ZEB2	DEG in HIV	705439	9839	zinc finger E-box binding homeobox 2	Rh02708878_m1	346

Acute SIV infection

Eight Indian-origin rhesus macaques were infected intravenously with 500 $TCID_{50}$ (50% tissue culture infective dose) of SIVmac239. Intravenous infection with a high dose of virus ensures a reliable and synchronous take of infection with a well-defined disease course. Four animals were sacrificed at day seven post infection and four animals were sacrificed at day 10 post infection. From previous studies, day 10 post infection is the known peak of plasma viral load and SIV gag DNA copies in CD4 memory T cells and was selected to maximize the chance

of detecting individual infected cells [163,165]. Lymphocytes were isolated from jejunum tissue and viably frozen.

Single cell qRT-PCR

Frozen viable lymphocytes isolated from the jejunums of each of the day 10 post infection animals were thawed, stained with antibodies to identify live CD4+ CD95+ memory T cells, and sorted using FACS. Using the Fluidigm C1 Single Cell Autoprep system, approximately 2000 cells in suspension were flowed through a microfluidic integrated fluidic circuit (IFC) with 96 single cell capture sites. Cells were stained with a viability dye and each capture site was visualized with a fluorescent microscope to verify the number and viability of captured cells. This system then lyses the cells, performs reverse transcription and gene-specific pre-amplification, and produces cDNA for each captured cell in nanoliter-scale reaction chambers. Output cDNA is then loaded into a Fluidigm BioMark IFC for qRT-PCR of 96 assays from each captured cell.

Jejunum CD4 memory T cells from each of the four day 10 animals were analyzed using a total of five C1 capture cycles. Of the possible 96 capture sites, 78 were filled on average for a total of 390 cells. A multi-step quality control procedure was used to ensure that each cell represented a single live cell. While captured in the C1 chip, cells were stained with live/dead dye and manually inspected using a microscope. Wells with more than one cell captured were noted and eliminated from downstream analysis. The live/dead status of each cell was also recorded. After gene expression analysis, the number of detectable genes expressed by each cell was determined, showing clear separation of three populations of cells **(Supplemental Figure 3.1)**. Empty wells expressing no detectable target genes were eliminated from further analysis.

Occasionally, a cell that stained as live failed to have any detectable assays expressed and was subsequently classified as an empty well and removed from analysis (8/62 empty wells = 13%). This could be due to a technical failure of the RNA isolation or cDNA conversion. A second group of cells expressed between four and 20 genes. These cells stained dead or had been identified as an empty well 87% of the time and were also removed from downstream analysis. A third group of cells expressed more than 20 genes. These cells stained live 90% of the time and were considered live, single cells. After quality control metrics were applied, there were on average 66 live cells per run for a total of 330 single cells analyzed.

Identification of SIV-infected T cells by single cell transcriptional profiling

Overall, at least one of the employed SIV RT-PCR assays yielded detectable signal in 24.5% of the cells. To ensure comparison of truly infected and uninfected cells, we eliminated cells expressing only one of the detected SIV transcripts from analysis resulting in 19.9% (61 of 307) productively SIV-infected live cells. This percentage of SIV-infected jejunal T cells in acute SIV infection is in agreement with previously published studies and suggests that the freeze/thaw and capture process did not significantly alter the proportion of infected cells [165,180]. As expected, the spliced Tat-Rev transcript was expressed in the lowest proportion of infected cells (25%), while the LTR, R-region, and Gag transcripts were expressed in a higher proportion of infected cells (80-95%).

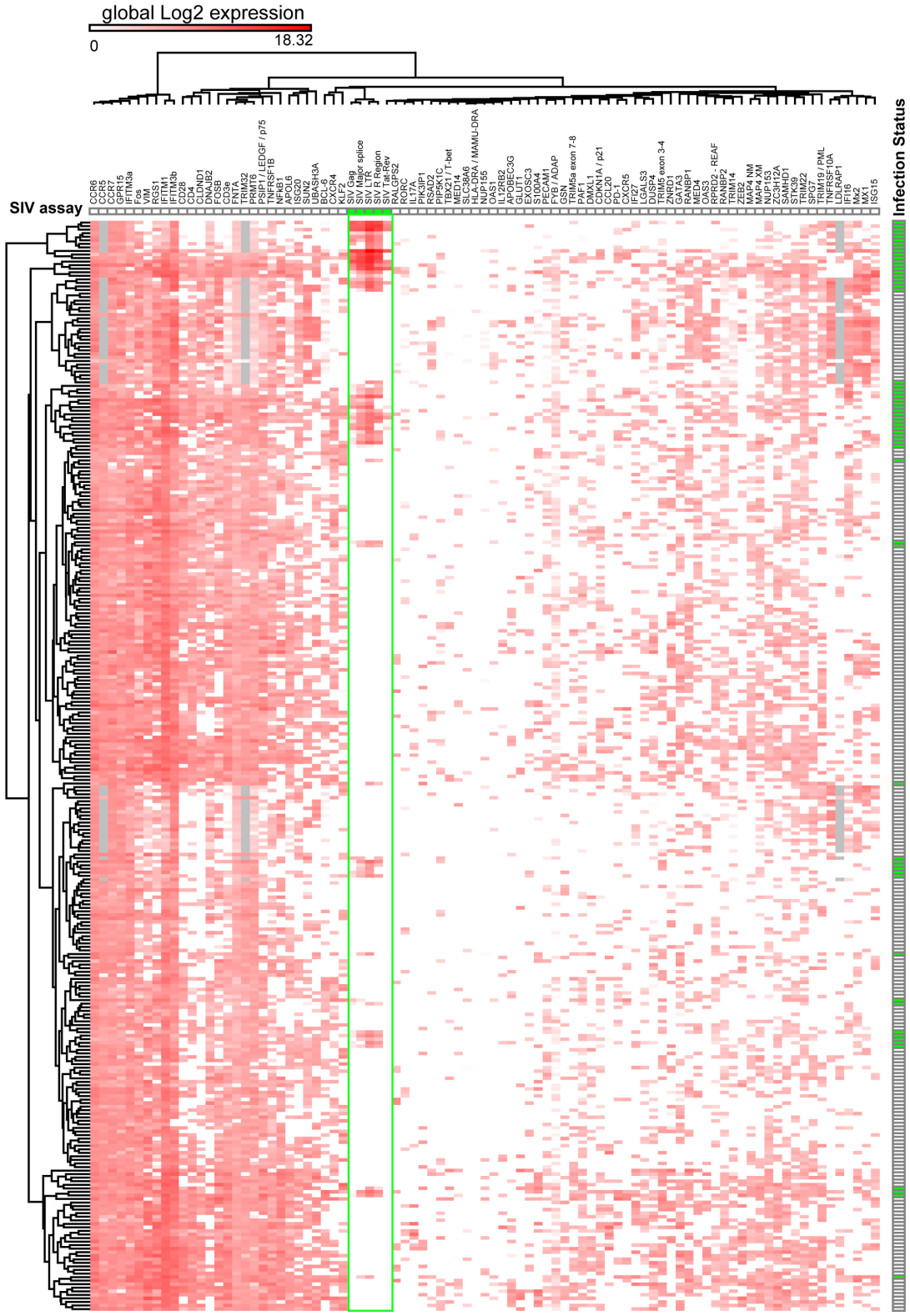
Analysis of single cell gene expression

Expression of all 89 genes (excluding endogenous control genes) in each of the 307 uninfected and infected single cells was visualized and cells clustered by similarity using

unsupervised hierarchical clustering **(Figure 3.1)**. This analysis places many of the SIV-infected cells in close proximity to each other, likely driven by their expression of SIV transcripts. The heatmap also shows the expected highly stochastic nature of single cell gene expression [280,362]. A complementary method of assessing similarity between cells is principal component analysis (PCA), a method to reduce the complexity of a high-dimension data set to the most informative few dimensions. Via PCA, all 307 cells were clustered based on expression of the 89 genes and colored by infection status, clearly demonstrating infected cells clustering separately from uninfected cells **(Figure 3.2.A.** and **B.)**. The loading factors for each axis are also shown **(Figure 3.2. C.-E.)** and the SIV assays are major drivers of the separation on all three axis.

Figure 3.1. Single cell gene expression from acute SIV infected jejunual CD4+ memory T cells identifies individual infected and uninfected cells.
Heatmap displaying Log2 expression data from 89 genes arrayed horizontally in 307 single cells arrayed vertically. Genes and cells were clustered using unsupervised hierarchical clustering with Euclidean distance and complete linkage parameters. SIV-specific assays are outlined in green and individual infected cells as defined by at least two SIV assays detected are highlighted in green. Grey represents missing data.

Figure 3.1. (Continued)



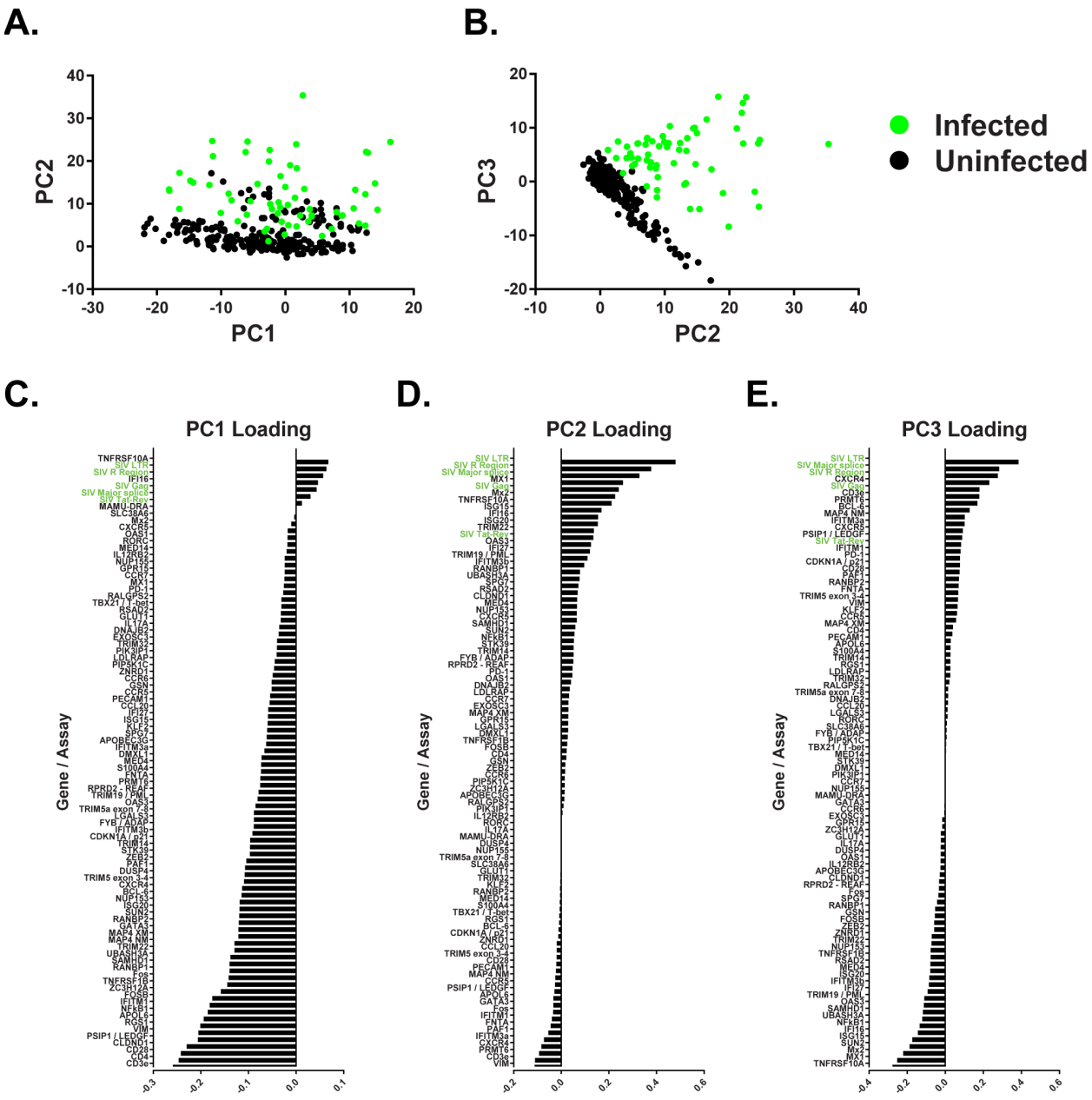


Figure 3.2. Principal component analysis clusters single cells by infection status. Principal component analysis (PCA) was performed based on all 89 genes/assays for 307 single cells **(A** and **B)**. Each symbol represents an individual cell and is colored based on infection status. Loading factors for PC1 **(C)** and PC2 **(D)** and PC3 **(E)** are shown with SIV assays highlighted in green.

Gene expression differences between infected and uninfected cells

Because not all genes are expressed in all single cells at all times, analysis of differential gene expression at the single cell level requires distinct approaches and methods as compared to analysis of gene expression in bulk populations of cells. Expression of genes in single cells often occurs in transient bursts resulting in stochastic effects that can result in the failure to detect expression of a gene at the time of analysis [280,362]. Furthermore, despite technical advances in the analysis of gene expression in single cells that provide a sensitivity of 1-5 molecules per cell [363,364], technical limitations of single cell analysis may result in failure to detect target RNAs. The net result of both of these factors is zero-inflated data sets, which require distinct approaches to deal with the two data components: discrete expression (i.e. the gene is on or off) and continuous expression (i.e. for expressed genes, the level of expression).

Due to zero-inflated data, traditional statistical methods such as t-tests fail to fully account for differences between two groups of single cell data sets. Since normal statistical distributional assumptions are violated, regular Poisson or negative binomial regression models cannot fit the data effectively. To find which individual genes are the most differentially expressed, we implemented a zero-inflated Poisson (ZIP) regression approach. Using the ZIP model, the top significant differentially expressed genes were CXCR5 (joint p value=0.00083), PD-1 (0.0046), Fos (0.011), APOL6 (0.021), and VIM (0.035) **(Figure 3.3.A)**. As a confirmation, a very similar published method [365] differing only in choice of joint test (a combined likelihood-ratio test) was used, and the results largely agreed, ordering the top four genes in the same rank order.

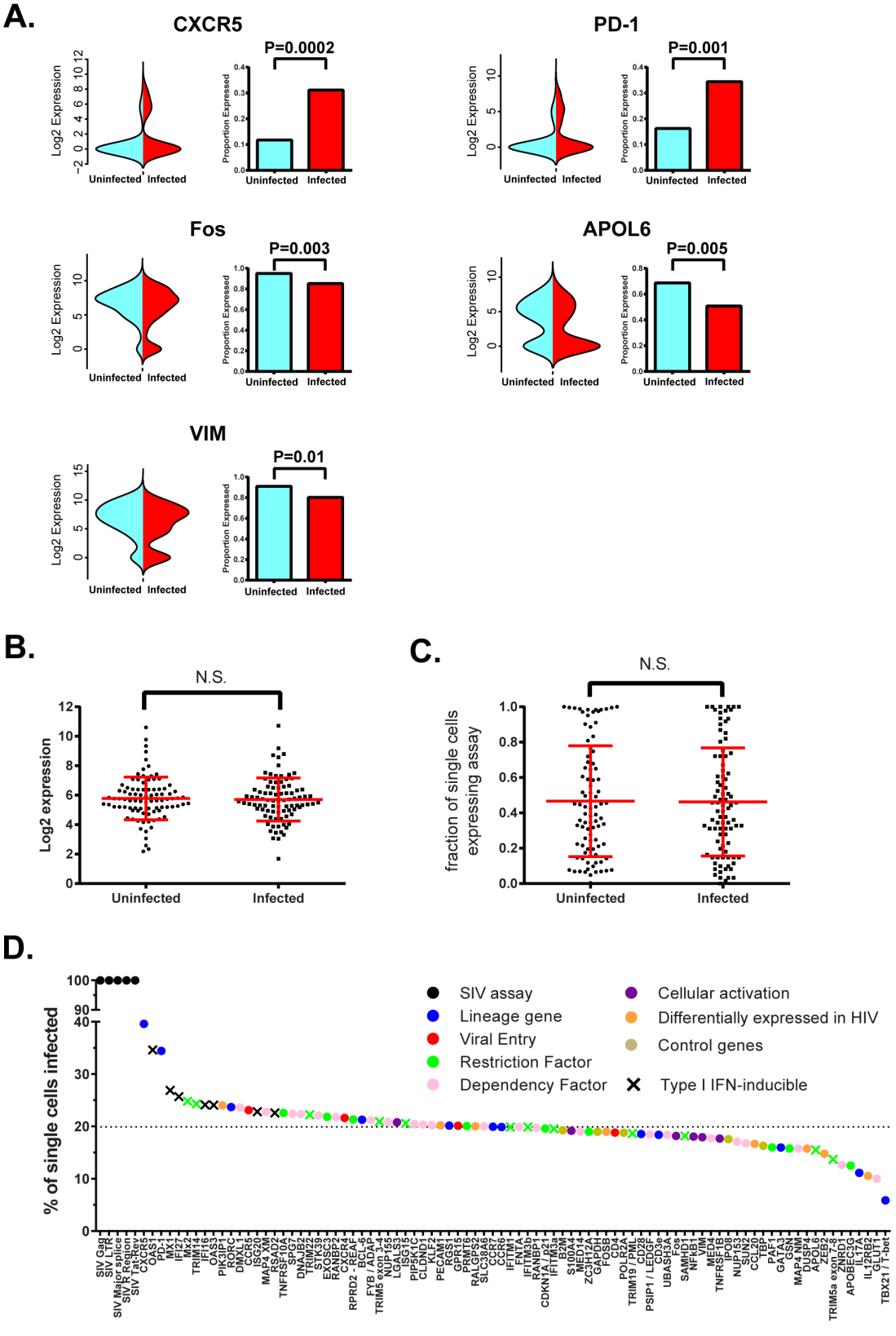
Importantly, uninfected and infected cells did not differ in continuous **(Figure 3.3.B)** or discrete **(Figure 3.3.C)** expression when all genes in the panel were aggregated. This suggests that cells which become infected are not more globally transcriptionally active.

A complementary method to identify differences between infected and uninfected cells is to assess the percent of cells that are infected if they express a certain gene **(Figure 3.3.D)**. For example, of the 177 cells expressing detectable CD4 transcripts, 18.8% were infected. The mean level of infection for cells expressing any individual cellular gene was 19.6%. Cells that expressed CXCR5 were 39.6% infected (the highest of any cellular gene) and cells that expressed PD-1 were 34.4% infected (third highest of any cellular gene). Rate ratios were calculated for each gene and a negative binomial model was used to calculate 95% confidence intervals for outlier analysis. Only CXCR5 and PD-1 were identified as outliers by this method.

Figure 3.3. Five genes were significantly differentially expressed in infected as compared to uninfected cells using zero-inflated Poisson statistics.

Expression of the five differentially expressed genes is shown in uninfected and infected cells **(A)**. The split violin plot is a smoothed representation of the detected Log2 expression level of the individual cells on the y-axis. The left side in blue is comprised of only uninfected cells (zero SIV assays detectable), while in red on the right side are infected cells (at least two SIV assays detectable). The bar chart displays the proportion of cells in each population that expressed detectable levels of the gene. P-values are the odds-ratio for detected expression (non-zero), infected vs. uninfected. Overall gene expression did not differ between infected and uninfected cells as measured by continuous **(B)** or discrete expression **(C)** using paired T-tests. The proportion of cells expressing each gene that was infected is displayed **(D)**. Outlier analysis identified CXCR5 and PD-1 as significant (after removal of SIV assays). Colors represent the category of gene, and an X symbol represents a gene that has been reported to be type I interferon-inducible.

Figure 3.3. (Continued)



Characterization of a novel cell population discovered through single cell expression analysis

Both statistical approaches described above identified CXCR5 and PD-1 as highly differentially expressed genes between infected and uninfected cells. Together CXCR5 and PD-1 are markers for T follicular helper cells (Tfh), [152] and both genes were enriched in the proportion of infected cells expressing them as compared to uninfected cells. Additionally, CXCR5 and PD-1 were coexpressed in the same uninfected single cell 3.7% of the time compared to 19.7% in infected cells, a significant enrichment as assessed by Fisher's exact test (p=0.0006).

These markers, especially if expressed in combination, suggest the presence of a distinct subpopulation of CD4 T cells that is infected at a high rate in acute SIV infection. The presence of Tfh cells in jejunum has not been previously described, however. To determine the presence, frequency, and cell surface phenotype of cells expressing CXCR5 and PD-1 in cells derived from jejunum, we used a 14-color flow cytometry panel **(Figure 3.4.A)**. The existence of a CXCR5+ PD-1+ population of memory CD4+ T cells in the jejunum of uninfected animals was confirmed. The population is similar to that seen in spleen or lymph node although lower in frequency, and this population is absent from peripheral blood **(Figure 3.4.B)**.

After verifying the existence of a PD-1+ CXCR5+ population of memory CD4+ T cells in jejunum, phenotypic markers characteristic of Tfh cells were analyzed in comparison to the same population from spleen or lymph node. Jejunum, spleen, and lymph node PD-1+ CXCR5+ cells expressed comparable levels of ICOS (CD278) and CD69 **(Figure 3.4.C)**. The level of both markers was higher than that in the bulk CD95+ memory CD4+ T cells from each compartment. Both proteins are known to be enriched in Tfh cells [366,367]. Jejunum and spleen PD-1+ CXCR5+ cells also expressed similar levels of a characteristic Tfh marker CD200 as well as CTLA4

(Figure 3.4.C), again at higher levels than the bulk CD95+ CD4+ T cells. CD200 is a marker of Tfh cells that is very specific to these cells [271].

Since CCR5 is the primary coreceptor for SIVmac and most transmitted HIV-1 strains [29,30,173], the level of expression of CCR5 could reflect the susceptibility of individual cells to infection. Similar to PD-1+ CXCR5+ T cells from spleen and lymph node, the PD-1+ CXCR5+ population in jejunum expressed low levels of CCR5 **(Figure 3.4.D)**. The PD-1^{mid} $CXCR5^{neg}$ population expressed the highest CCR5 levels in each tissue of the tested populations.

The only phenotypic difference found between jejunum and spleen PD-1+ CXCR5+ cells was increased CCR7 expression in jejunum. This was despite generally lower CCR7 expression in jejunum total CD4+ T cells **(Figure 3.4.E)**, consistent with lower expression of CCR7 on T cells from extralymphoid locations compared with the spleen.

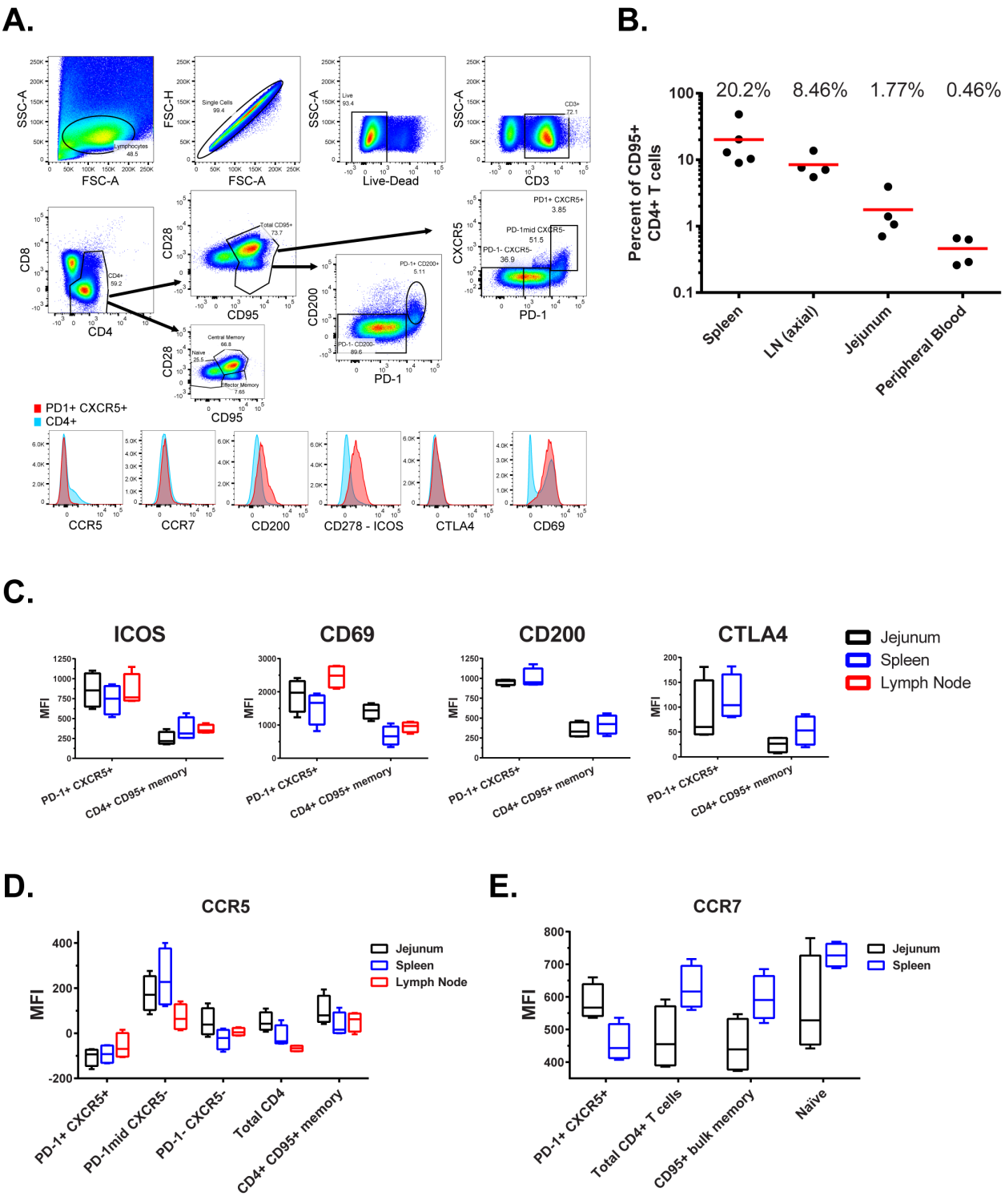


Figure 3.4. PD-1+ CXCR5+ cells from SIV-uninfected jejunum exhibit phenotypic characteristics of Tfh cells.

The gating strategy for phenotypic analysis of PD-1 and CXCR5 cell populations and example histograms showing expression of phenotypic markers in the PD-1+ CXCR5+ population as compared to total CD4+ T cells is shown **(A)**. The frequency of PD-1+ CXCR5+ cells as a

Figure 3.4. (Continued)
percentage of total CD95+ memory CD4+ T cells in various anatomical locations is shown **(B)**. The numbers above each location represent the mean of 4-5 animals. The MFI of surface markers typically found on Tfh cells is shown for cells from jejunum, spleen, and axial lymph node from four animals **(C)**. The MFIs of CCR5 in defined populations of CD4+ T cells from jejunum, spleen, and lymph node are shown **(D)**. The MFIs of CCR7 in defined populations of CD4+ T cells from jejunum and spleen are shown **(E)**.

A defining characteristic of Tfh cells is the expression of the transcription factor BCL6 [257]. Using intracellular staining and flow cytometry, a similar level of BCL6 protein expression was found in jejunum PD-1+ CXCR5+ cells as in the corresponding population in spleen and lymph node **(Figure 3.5.A)**. The primary cytokine produced by Tfh cells critical for their function is IL-21 [152]. Production of IL-21 mRNA is significantly higher in sorted PD-1+ CXCR5+ cells compared to other sorted subsets **(Figure 3.5.B)**.

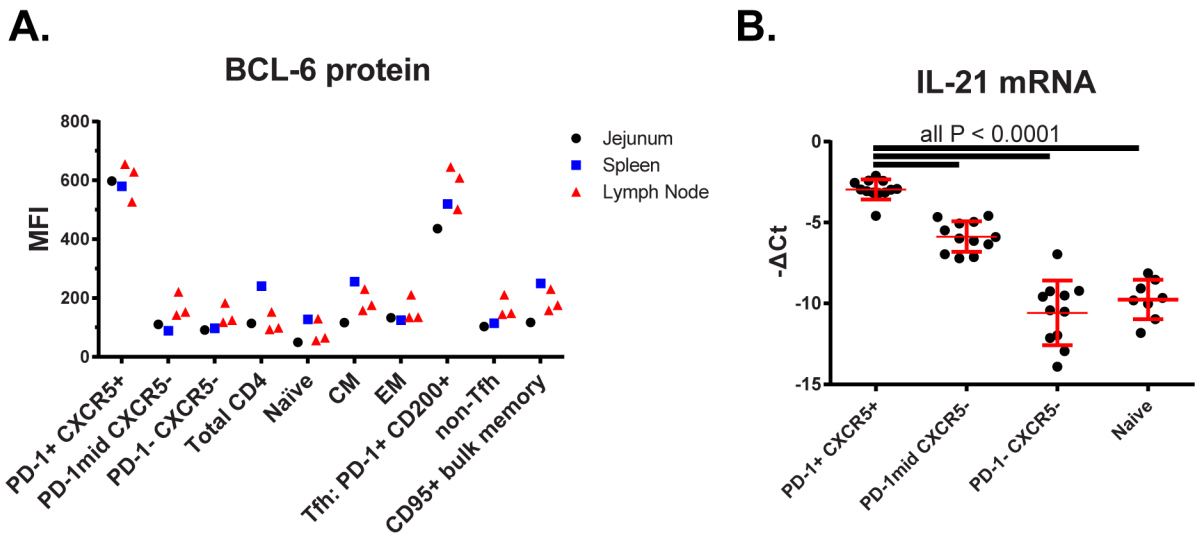


Figure 3.5. Functional characterization of PD-1+ CXCR5+ cells shows Tfh characteristics. Intracellular staining for BCL6 protein in jejunum, spleen, and lymph node populations is displayed as mean fluorescent intensity **(A)**. Each point represents an individual sample. The expression of IL-21 mRNA in sorted populations of PD-1 and CXCR5 cells compared to CD95- $CD28^{mid}$ naïve CD4+ T cells **(B)**. $p < 0.0001$ by paired T test for each comparison with the PD-1+ CXCR5+ population.

High-throughput gene expression profiling

Analysis of Tfh phenotypic markers and BCL-6 protein expression suggested that the PD-1+ CXCR5+ memory CD4+ T cell population from jejunum was similar to that of Tfh cells found in spleen or lymph node. To better assess the potential similarity of these populations, a panel of 77 Tfh-associated genes was identified. Genes were selected that were known to be enriched in Tfh cells such as *IL-4* and *IL-21*, or decreased in Tfh cells such as *IL-4R* and *GATA3*. Using RNA-Seq data from a recent publication [276], template matching [368] was used to discover additional, less well known, Tfh genes. We utilized the published RNA-Seq data from four sorted populations of cells: bulk CD4+ T cells, Tfh cells (CXCR5+PD1^{hi}CD127^{-}CD25^{-}), T follicular regulatory cells (TfhR) (CXCR5^{+}PD1^{hi}CD127^{-}CD25^{+}), and T regulatory cells (CXCR5$^{+/-}$PD1$^{lo/int}$CD127^{-} CD25^{+}). Template matching is an algorithm that ranks the expression profile of each gene for its correlation to a predefined template. In this case, we sought to find genes whose expression was high in Tfh cells while low in the other three cell populations, or high in Tfh and TfhR cells while low in the other two. As for any correlation measure, the output is an R- and P-value for each gene. The entire list of over 10,000 expressed genes can be ranked by R-value, ordering the genes that best match the template (for example, high expression in Tfh and low in other cells) with the highest R-value to the genes that least match the template with the lowest R-value. This allowed for an unbiased selection of the most differentially expressed genes between Tfh cells and other CD4+ T cells. Markers classically associated with Tfh cells such as CD200, PD-1, and BCL6 all ranked in the top 30 genes, validating the method.

To analyze the similarity of gene expression in jejunum PD-1+ CXCR5+ cells and classical Tfh cells, the following PD-1 and CXCR5 memory CD4+ T cell populations were sorted from jejunum and axial lymph node: PD-1+ CXCR5+, PD-1^{mid} CXCR5-, and PD-1-

CXCR5-, as well as CD95- CD28mid naïve CD4+ T cells. In addition, cells using a common definition for Tfh, memory CD4+ T cells expressing PD-1 and CD200, were sorted, and gene expression was measured for all 77 genes in the panel. Expression was normalized to *ACTB*, the most stable control gene as determined by the NormFinder algorithm [319]. Using unsupervised hierarchical clustering, a clear group of PD-1+ CXCR5+ cells from jejunum, PD-1+ CXCR5+ cells from axial lymph node, and PD-1+ CD200+ cells from lymph node was found **(Figure 3.6)** demonstrating their similar expression of over 77 Tfh-related genes. Naïve CD4+ T cells also segregated together, while the PD-1mid CXCR5- and PD-1- CXCR5- populations showed intermediate expression of many Tfh-related genes. Some differences were attributable to the tissue of origin. For example, lower *CCR7* and *KLF2* expression in lymph node cells. However, these differences were smaller than those between the sorted populations. SIV infection status of the animals did not affect the overall clustering of cells by PD-1 and CXCR5 surface phenotype since samples from infected and uninfected animals were interspersed in the major clusters.

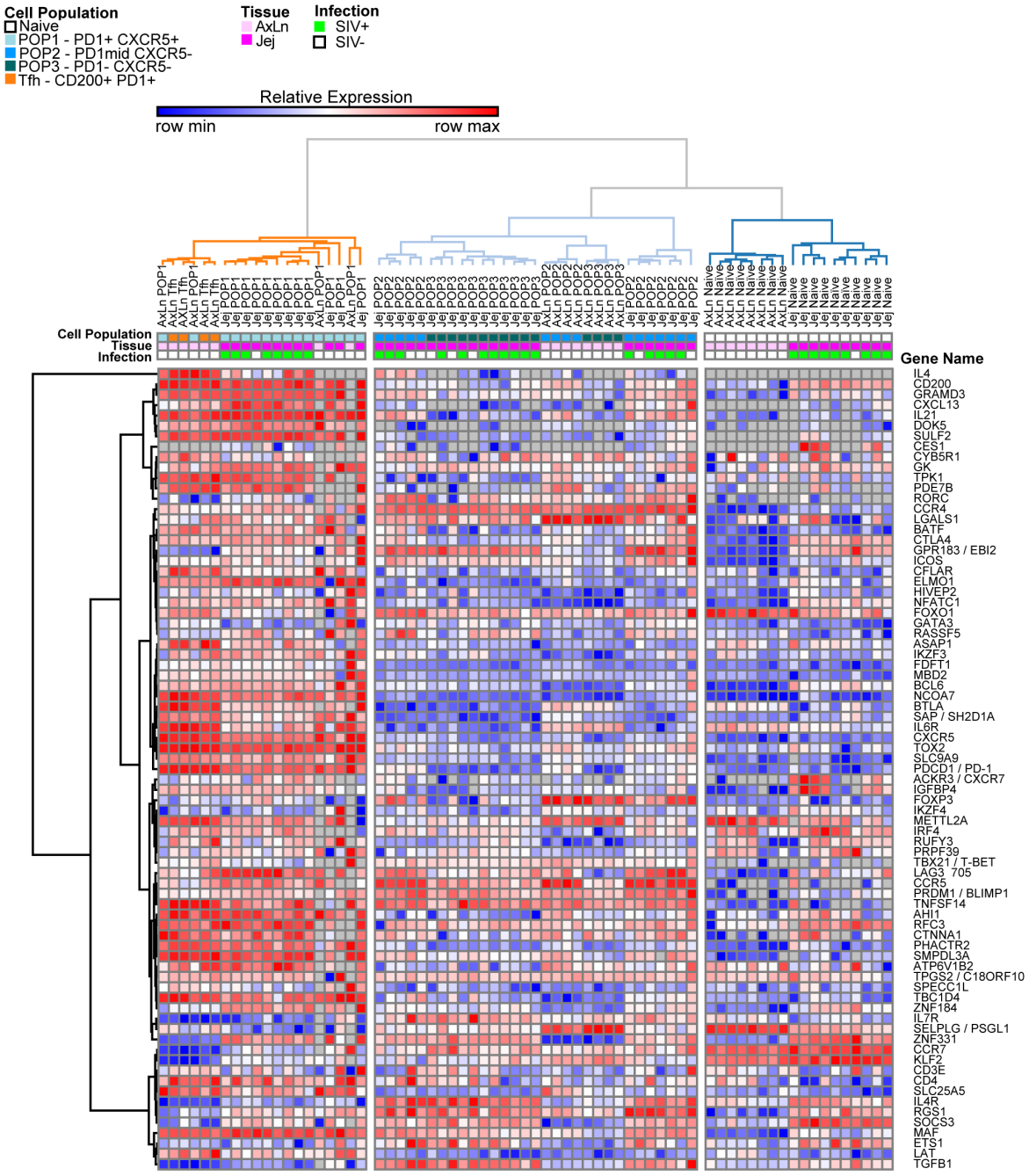


Figure 3.6. The expression profile of PD-1+ CXCR5+ cells from jejunum clusters with that of Tfh cells from lymph nodes.
Sorted subpopulations of PD-1/CXCR5 memory CD4+ T cells as well as naïve cells were examined for expression of a panel of 77 genes that exhibited significant differential expression in Tfh cells as compared to bulk CD4+ T cells or Treg cells. Normalized results are displayed using unsupervised hierarchical clustering. Samples are coded for cell population, tissue of origin, and infection status using color bars under the sample names.

Principal component analysis with all 77 genes also tightly grouped PD-1+ CXCR5+ cells from jejunum with PD-1+ CXCR5+ cells from axial lymph node and PD-1+ CD200+ defined Tfh cells from axial lymph node **(Figure 3.7.A)**. As shown by the loading factors, the clear separation on the PC1 axis is driven by genes known to be enriched in Tfh cells such as *IL21*, as well as genes known to be decreased in Tfh cells, such as *IL4R* and *IL7R*, thus demonstrating the Tfh-like characteristics of the PD-1+ CXCR5+ population **(Figure 3.7.B)**. Importantly, clustering was primarily based on cell subset, not tissue location or infection status of the cells. The tight similarity of the PD-1+ CXCR5+ cells to Tfh cells from lymph nodes in the expression of 77 genes strongly suggests these cells are also Tfh cells.

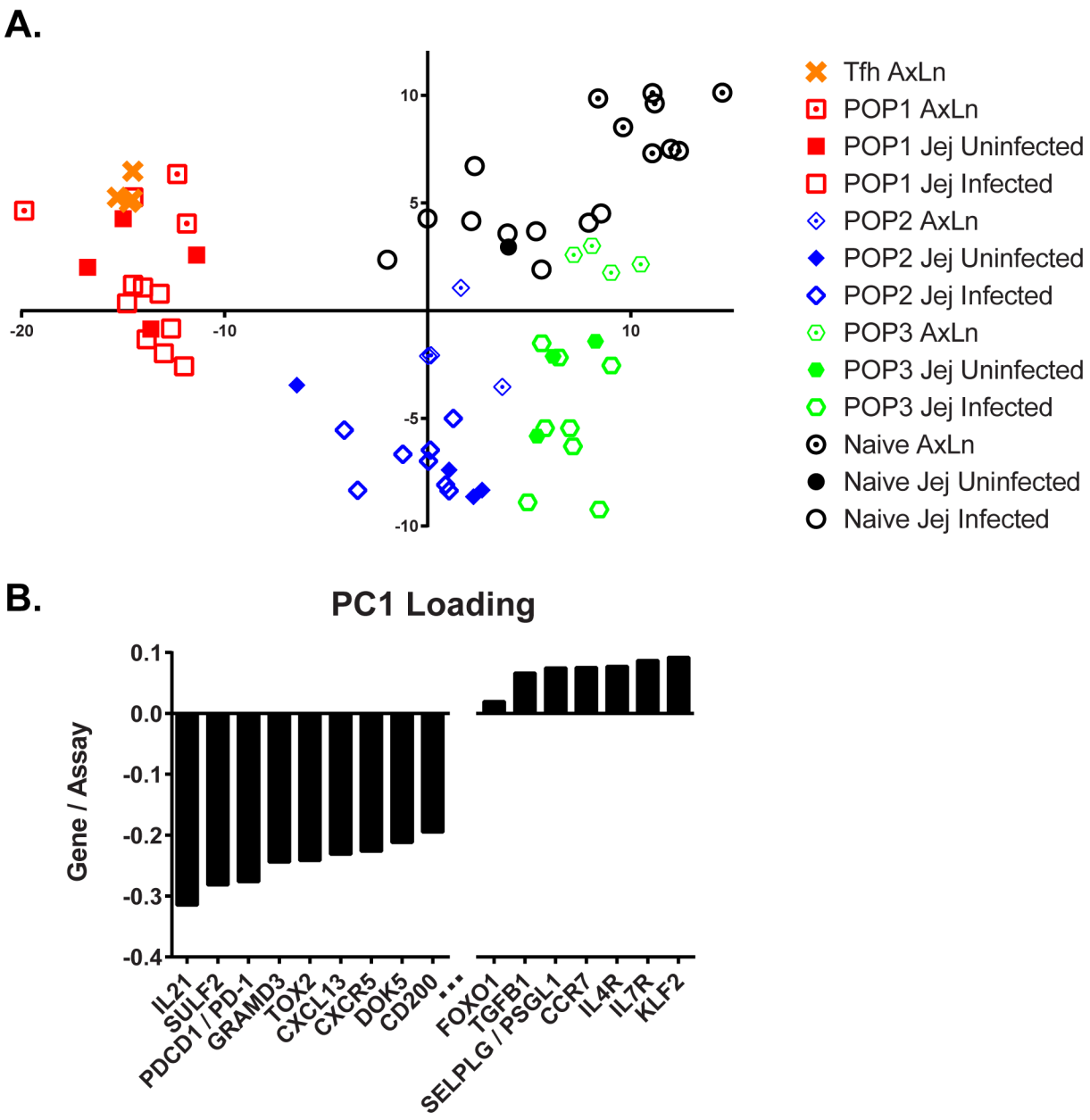


Figure 3.7. Principal component analysis clusters PD-1+ CXCR5+ cells from jejunum with Tfh cells.
Using the panel of 77 Tfh-related genes, PCA was used to cluster each sorted subset of PD-1 and CXCR5 memory CD4+ T cells or naïve cells from both axial lymph nodes and jejunum **(A)**. POP1 = PD-1+ CXCR5+, POP2 = PD-1mid CXCR5-, POP3 = PD-1- CXCR5-, and naïve = CD95- CD28mid. The genes that were the strongest drivers of differentiation on the PC1 axis are displayed as loading factors **(B)**. They included genes with known high expression in Tfh cells like *IL21* and *CD200* as well as genes with known low expression in Tfh cells like *CCR7* and *IL7R*.

Acute SIV infection reveals high susceptibility of PD-1+ CXCR5+ cells

Populations of CD4+ T cells were sorted from jejunum of SIV infected rhesus macaques at either seven or 10 days post infection. As before, the sorted populations were naïve (CD28mid CD95+) and memory CD95+ cells from PD-1+ CXCR5+, PD-1mid CXCR5-, and PD-1- CXCR5- subpopulations. After DNA/RNA extraction, the levels of SIV Gag DNA and RNA were quantified in comparison to standard curves of known concentrations **(Figure 3.8A** and **B)**. At 10 days post infection, the PD-1+ CXCR5+ population had the highest level of SIV Gag DNA (mean of 3.41 copies/cell) and RNA (mean of 228 copies/cell), substantially more than any other population. Strikingly, the PD-1+ CXCR5+ population is more highly infected than the PD-1mid CXCR5- population which expressed higher levels of the SIV coreceptor CCR5 surface protein **(Figure 3.4.D)** and mRNA **(Figure 3.6)**. Levels of SIV Gag DNA and RNA were much lower for animals at seven days post infection, with only one of the four animals having detectable virus in all four cell populations (data not shown).

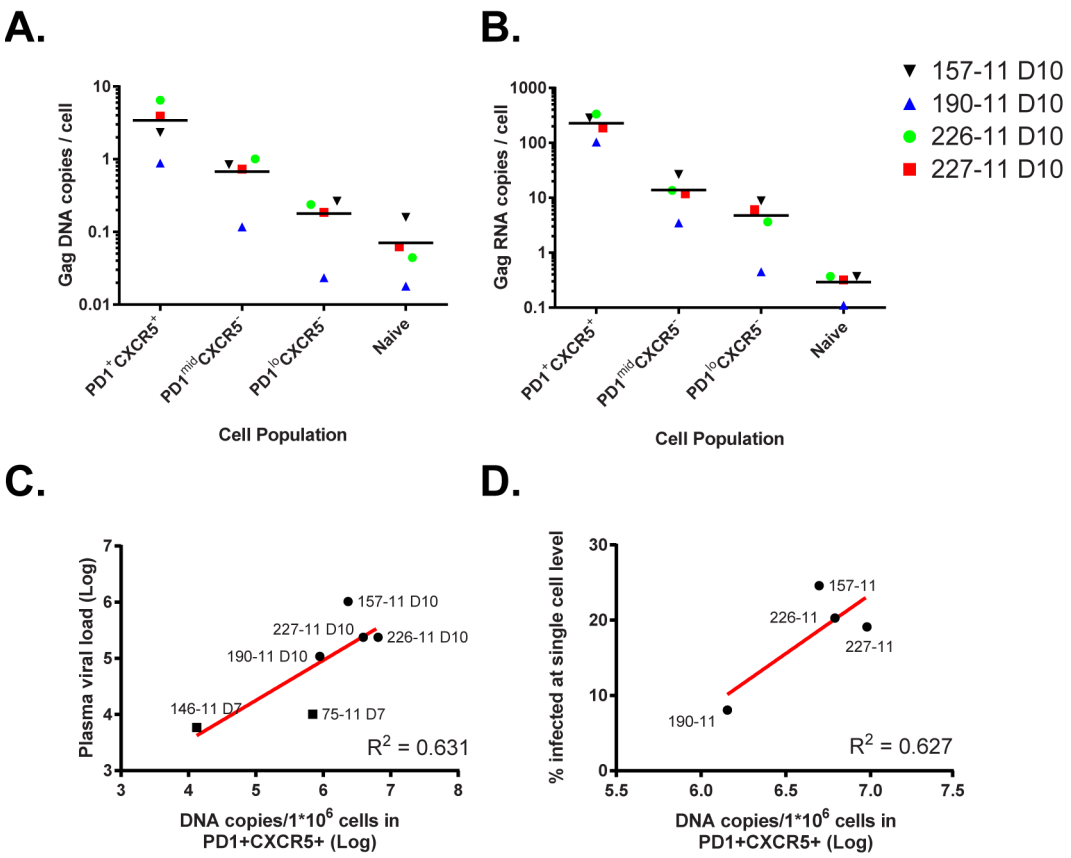


Figure 3.8. PD-1+ CXCR5+ cells from jejunum are highly susceptible to acute SIV infection.
Memory CD4+ T cells were sorted and the *Gag* DNA copies per cell were quantified by qRT-PCR with a standard curve **(A)**. Gag RNA copies from the same cell populations were also quantified **(B)**. Plasma viral load **(C)** and the percent of single cells that were infected **(D)** show positive relationships with the total *Gag* DNA copies from the PD-1+ CXCR5+ memory CD4+ T cell population.

Plasma viral loads were measured for all eight animals using quantitative RT-PCR [354,355]. When compared to the animals with detectable SIV *gag* DNA in the PD-1+ CXCR5+ jejunal cells, a positive relationship emerged, although it did not reach significance (p=0.07, Spearman correlation), likely due to the relatively small sample size. This suggests that plasma viral load is reflective of tissue viral burden. Similarly, a non-statistically significant positive relationship was found between the SIV *gag* DNA copies per cell in the PD-1+ CXCR5+ population and the

percent of cells that became infected at a single cell level in the day 10 post infection animals. This suggests that the single cell capture and processing did not bias selection of infected cells.

Overall, single cell gene expression measurements identified a subpopulation of memory CD4+ T cells that are highly susceptible to acute SIV infection *in vivo*. These cells were further characterized to have many phenotypic, transcriptional, and functional features of Tfh cells despite their localization in jejunum tissue, where Tfh cells have not been previously described.

3.E. Discussion

A complete molecular description of an *in vivo* target cell for immunodeficiency virus infection is currently lacking. Traditional gene expression measurements that average expression from multiple cells may overlook critical subpopulations. Through the use of single cell qRT-PCR, memory CD4+ T cells from acute *in vivo* SIV-infected rhesus macaque mucosal tissue were analyzed for the expression of genes predicted to distinguish why certain cells became infected. Individual infected and uninfected cells were identified through qRT-PCR of multiple SIV sequences, and the level of infection averaged 20%. Lineage markers PD-1 and CXCR5, with nearly 20% of infected cells expressing both genes compared to just 3.7% of uninfected cells, were the most differentially expressed genes. These cell surface proteins can define Tfh cells; however, existence of a PD-1+ CXCR5+ population in jejunum had not previously been described. A reproducible population with a frequency intermediate to that in peripheral blood and lymph node was found to exist in jejunum and expressed cell surface molecules characteristic of Tfh such as ICOS, CD69, CTLA4, CCR5, and CD200. In addition, the jejunal PD-1 CXCR5+ population expressed the canonical Tfh transcription factor BCL6 at similar levels to those observed in classical Tfh populations from lymph node or spleen. Additionally,

the expression profile of over 70 Tfh-associated genes in the PD-1+ CXCR5+ cells from jejunum was remarkably similar to classical Tfh cells. Finally, the PD-1+ CXCR5+ cells from jejunum were found to be highly SIV-infected, with an average of over three SIV DNA copies/cell at 10 days post infection.

Gene expression at the single cell level, even among phenotypically identical cells, can vary widely due to both intrinsic (reactions involving small numbers of molecules are dominated by large concentration fluctuations and inherently noisy) [369] and extrinsic (microenvironment) factors [277]. Overcoming noise at the single cell level is possible by analyzing relatively large numbers of genes in tens to hundreds of cells, a process that is facilitated by instrumentation for high-throughput transcriptional profiling [278]. Using single cell qRT-PCR, it is possible to dissect heterogeneous tissues into cell subpopulations according to their unique gene expression profiles and uncover rare cell populations that may be hidden in bulk analysis [370-372]. Additionally, studies at the single cell level have shown that the noise or variation in protein levels is primarily predicted by variation in mRNA levels [373-375]. Therefore, measuring mRNA is an accurate proxy for protein levels in most cases, even in single cells.

Very few studies have measured viral infection at the single cell level [376], and fewer have measured cellular gene expression in single infected and uninfected cells. The only *in vivo* single cell viral infection study published to date used rotavirus to infect mice and measured gene expression in purified, sorted single intestinal epithelial cells (IECs) by qRT-PCR [377]. This method accounts for possible differences between directly infected cells and bystander cells, while bulk analysis is generally unable to distinguish the two (for example, comparing bulk cells from infected and uninfected animals). Although the population of IECs was homogenous based on surface marker expression, considerable heterogeneity was found in response to virus in both

virus-infected and bystander cells. The authors speculated that different physical position along the crypt-villus and proximal-distal axes plus distance from regulatory non-epithelial cells such as dendritic cells could result in different microenvironments and therefore different transcriptional responses to infection. The pre-sorting of cells for populations infected at higher (CD26+, 200-fold increased infection) and lower rates (CD44+ crypt cells, resistant to infection), while necessary for obtaining sufficient virally infected cells, did not allow for determining differences in each subpopulation's susceptibility to infection or the frequency of infection. Additionally, the study focused on the subversion of the interferon response by rotavirus and accordingly selected a panel of genes related to interferon. Here we took advantage of the very high level of infection within mucosal memory CD4+ T cells during acute SIV infection [165] to avoid biasing the cell selection and used a panel of genes predicted to influence cell susceptibility to infection. This approach allowed for discovery of an undescribed subset of cells harboring high levels of virus.

Classical T follicular helper cells (Tfh) are a CD4+ helper T cell subset specialized for providing help for B cells [152]. They are essential for germinal center formation, affinity maturation, and class-switch recombination, and allow the formation of antibody-producing plasma B cells and long-lived memory B cells via their expression of CD40 ligand, IL-21, IL-4 and other molecules. ICOS expression is important for Tfh differentiation, migration into follicles, and functions [248,253,378]. Similar levels of ICOS expression were seen in the PD-1+ CXCR5+ cells from jejunum with spleen and lymph node Tfh cells in this study. Tfh differentiation requires expression of the transcription factor BCL6 [251,256,379]. BCL6 expression likely promotes Tfh differentiation by multiple mechanisms that have not been fully described. However, it is known to repress Blimp-1 [256] as well as Th1, Th2, Th17, and Treg transcription

factors, and cytokine genes [251,257]. Here we demonstrated that BCL6 protein and transcript production in PD-1+ CXCR5+ cells from jejunum reaches similar levels as Tfh cells from lymph node and spleen.

The first reports characterizing Tfh cells were in tonsil lymphoid tissue [366,380], and much of the Tfh research to date has focused on secondary lymphoid tissue, particularly those in germinal centers. More recently, Tfh-like cells have been found in peripheral blood that express CXCR5 and share functional properties with Tfh cells, including secretion of IL-21 and induction of naïve B cell proliferation, differentiation, and class-switching (reviewed in [381,382]). These circulating Tfh cells do not express BCL6 protein and express relatively low levels of activation markers like ICOS and CD69. Subsets of circulating Tfh cells do express the cannonical transcription factors for Th1 (T-bet), Th2 (Gata3), and Th17 (RORγt) cells. Here we show that the PD-1+ CXCR5+ cells from jejunum express BCL6 protein and activation markers at comparable levels to Tfh cells from lymph nodes, and do not express T-bet, Gata3, or RORγt **(Figure 3.6)**, thus distinguishing this population from circulating Tfh cells.

Though similar to classical germinal center Tfh cells in many phenotypic and transcriptional respects, the jejunum PD-1+ CXCR5+ population described here differs by expressing higher levels of CCR7 **(Figure 3.4.E** and **Figure 3.6)**. A study found that CXCR5 overexpression was not sufficient to promote follicular entry of naive T cells unless the counterbalancing CCR7 expression was reduced [383]. This, along with the lack of organized lymphoid structures, suggests that the jejunum PD-1+ CXCR5+ cells do not reside in organized follicles. Additionally, localization of Tfh to follicles is aided by GPR183/EBI2 expression, while final localization to germinal centers requires reduction of GPR183 expression [384]. Here we show Tfh cells from lymph nodes expressed lower levels GPR183 compared to the jejunum

PD-1+ CXCR5+ cells **(Figure 3.6)**. A final indication that these cells are not likely to be located in germinal centers is their differential expression of SH2D1A/SAP compared to Tfh cells from lymph nodes **(Figure 3.6)**. In the complete absence of SAP, Tfh cells have defective adhesion to GC B cells and fail to be retained in GCs, and loss of SAP leads to complete loss of germinal center Tfh cells and germinal center B cells [385]. Here the jejunum PD-1+ CXCR5+ cells do express SAP, but at lower levels than Tfh cells from lymph nodes.

Peyer's patches, lymph nodes located primarily in the distal ileum [386], have been known to harbor Tfh cells [387,388]. Only limited recent evidence suggests Tfh or Tfh-like cells can exist outside of classically defined immune inductive sites in the mucosa. Mycobacterium tuberculosis-specific PD-1^{high} cells residing in the lung parenchyma required intrinsic Bcl6 and ICOS expression for their generation, and a fraction of these cells co-expressed CXCR5 [351]. In a humanized-DRAG mouse model, human Tfh cells were found in gut lamina propria at about 1/3 the frequency of that in Peyer's patches or mesenteric lymph nodes [352]. Here we describe Tfh-like cells in rhesus macaque jejunum, a site devoid of organized lymphoid tissue. The precise function of the jejunum PD-1+ CXCR5+ cells and whether they are tissue-resident or transiting between lymph nodes remain to be investigated.

It is well established that follicles are major sites of HIV [266-269] and SIV [266,270] replication and that Tfh cells are targets for HIV and SIV infection [265]. One mechanism of enhanced follicular replication could be the lack of CD8+ cytotoxic lymphocytes (CTLs) in follicles, as this concentrated replication does not exist in acute infection prior to the CTL response [270] or after CD8 depletion [271]. A previous study failed to find PD-1^{high} cells in jejunum during chronic SIV infection [350]. The high level of SIV DNA and RNA in the extra-follicular PD-1+ CXCR5+ population during acute infection shown here suggests these cells could be depleted early in SIV

infection. A recent study bolsters this view, having shown depletion of splenic Tfh cells in acute SIV infection of rhesus macaques [272]. Alternatively, if these cells migrate to lymph node follicles they could be protected from CTL-mediated killing and potentially seed a latent reservoir [271].

The PD-1+ CXCR5+ jejunum cells described here express lower levels of CCR5 than do the PD-1^{mid} CXCR5- population **(Figure 3.4.D),** yet are infected at a much higher rate **(Figure 3.8.A** and **B)**. The vast weight of evidence points to activated, memory CD4+ T cells that express CCR5 as the preferred cellular target for HIV and SIV infection [166,389,390]. In addition to infection studies, most transmitted founder viruses use CCR5 as a coreceptor [173], while CXCR4 usage generally only develops later in infection [32]. Like HIV-1, SIVmac primarily uses CCR5 as a coreceptor [30]. Although much evidence points to the importance of CCR5 in viral entry, it has been established that Tfh cells from germinal centers, which express low levels of CCR5, are found to be infected by HIV and SIV [265,391].

Some studies have described infection in memory CD4+ T cells that do not express very high surface levels of CCR5 yet are infected as measured by SIV DNA [165]. Our findings here are in agreement with Mattapallil et al.: while CCR5 surface protein expression was low and similar to naïve cells, CCR5 mRNA expression was greater in the PD-1+ CXCR5+ population than in naïve cells **(Figure 3.6)**. How these cells are infected despite low CCR5 surface expression remains to be discovered. Possibilities include a requirement for only a low level of CCR5 expression, or the cells becoming infected while expressing higher levels of CCR5 and later decreasing surface expression. Tfh precursor cells (CXCR5- CD4+ T cells) in lymph nodes express higher levels of CCR5 before entering germinal centers, and this has been proposed as the point at which Tfh cells become infected [391]. Alternative coreceptor usage is another possibility. However, in pigtail macaques, Tfh cells from lymph node and spleen were found to

express low levels of CCR5 and the alternative SIV coreceptors CXCR6 and GPR15, yet were found to be highly infected early in pathogenic SIV infection [392].

The level of infection in the PD-1+ CXCR5+ population is very high at 3.4 copies of SIV Gag DNA per cell on average. Mattapallil et al. measured 1.5 SIV Gag DNA copies per cell in bulk memory CD4+ T cells from rhesus macaque jejunum on day 10 post infection [165]. The highest level of infection in specific CD4+ T cell subsets reported to date during acute SIV infection showed a similar level of infection to what we observed in $\alpha4+\beta7^{high}$ memory T cells in peripheral blood at 10 days post infection in SIVmac251-infected rhesus macaques [340].

Overall, single cell gene expression analysis of individual infected and uninfected cells revealed a novel population of PD-1+ CXCR5+ memory CD4+ T cells in rhesus macaque jejunum that have many phenotypic and transcriptional similarities to Tfh cells.

Chapter 4: Conclusions and Discussion

4.A. Conclusions

The second chapter in this thesis measured gene expression of putative and confirmed restriction and dependency factor genes in defined subsets of CD4+ T cells from peripheral blood and jejunum of rhesus macaques. Since memory CD4+ T cells, and specifically those expressing CCR5, are particularly susceptible to HIV and SIV infection, we sorted CD4+ T cells into defined memory subsets, including CCR5+ cells, and found an overall increase in dependency factors in all memory subsets relative to naïve CD4+ T cells. The upregulation of genes that are required for SIV replication (dependency factors) in memory CD4+ T cells may contribute to the increased susceptibility of this population to SIV infection relative to naïve CD4+ T cells. Memory differentiation dynamically modulated expression of restriction factors resulting in both upregulation and downregulation of restriction factors. For both restriction and dependency factors the expression patterns of those groups of genes alone were sufficient to cluster memory subsets together using principal component analysis, demonstrating a reproducible pattern of expression due to memory differentiation. Stimulation of CD4+ T cells with either anti-CD3/CD28 or type I interferon also led to modulation of expression of restriction factors. However, the patterns of modulation following these stimulation modalities did not mirror those observed in the process of memory T cell differentiation.

CD4+ T cells from the intestinal mucosa are the primary target cells in acute SIV and HIV infection and are infected at relatively high levels and depleted rapidly, demonstrating that they are much more susceptible to infection than cells from peripheral blood. Prior to these studies, there was no information on differences in gene expression of restriction and dependency factors in CD4 T cells found in peripheral blood or mucosal sites, in large part because this gastrointestinal lymphoid tissue, especially in the small bowel, is difficult to sample

in humans. We measured gene expression in the same naïve and memory CD4+ T cell subsets from jejunum cells and found overall similar patterns due to memory differentiation, which also allowed clustering similar to peripheral blood. Despite the broad similarity, there was a significant decrease in the average expression of all restriction factors in the CCR5+ memory CD4+ T cells from jejunum relative to the same subset from peripheral blood. We suggest that this difference may correspond to their increased susceptibility to infection. Finally, acute infection with pathogenic SIVmac239 verified that the jejunum CCR5+ memory CD4+ T cells had the highest level of cell-associated viral RNA of all memory subsets in acute SIV infection, and that infection was associated with a broad increase in expression of nearly all restriction factors measured.

The third chapter in this thesis sought to measure potential gene expression differences at a single cell level in CD4+ T cells from the intestinal mucosa during acute SIV infection. Over 300 single cells were analyzed for the expression of 96 genes and the presence of SIV RNA transcripts indicating infection. Comparing the approximately 20% of T cells that were infected to uninfected T cells using multiple statistical methods revealed PD-1 and CXCR5 as the most differentially expressed genes, with both significantly enriched in infected cells. Expression of both PD-1 and CXCR5 can define Tfh cells, however the jejunum lacks the organized secondary lymphoid sites where classical germinal center Tfh cells are typically found. Using flow cytometry, we verified the existence of a PD-1+ CXCR5+ memory CD4+ T cell population in jejunum, which was observed at significantly higher frequencies than peripheral blood but at lower levels than that in spleen or lymph node. This population expressed cell surface markers typical of Tfh cells (ICOS, CD69, CTLA4, and CD200) as well as the canonical Tfh transcription factor BCL6 in similar proportion to Tfh cells from spleen or lymph nodes,

distinguishing it from less-activated circulating Tfh cells. To confirm the Tfh-like identity of the PD-1+ CXCR5+ population from jejunum, we developed a panel of 77 Tfh-associated genes and showed remarkable similarity in expression profiles between this subset and classical Tfh cells from lymph nodes. The PD-1+ CXCR5+ population from jejunum was found to be highly SIV-infected by cell-associated DNA and RNA, with an average of over three SIV DNA copies/cell at 10 days post infection.

4.B. Restriction and dependency factor gene expression

Combinatorial effects of restriction factors

While study of individual genes is critical for determining their relative contributions and mechanism(s) of restriction, the impact of restriction factors on viral replication is likely cumulative, since they act on different steps of the viral replication cycle to impair replication. Two prior studies have set out to undertake a comprehensive analysis of HIV restriction factors in CD4+ T cells. The first measured expression of 34 potential restriction factors in elite controllers as compared to untreated non-controls, cART suppressed, and uninfected individuals [285]. They found that the cumulative fold-difference in restriction factor expression with respect to a control individual was positively correlated with CD4+ T cell activation, viral load, and the expression of ISG15, which is a marker of interferon exposure. These data support our conclusions that memory CD4+ T cells, which are more activated than naïve, express generally higher levels of restriction factors. Drawbacks to this study include the measurement of restriction factor expression only in peripheral blood, which harbors very few infected cells in chronic infection, and the use of total CD4+ T cells. Here we showed significant differential modulation in naïve and memory subsets, demonstrating that analysis of bulk CD4+ T cells does

not accurately reflect the actual target cells for HIV infection. A second study from the same group measured expression of 32 restriction factors longitudinally in six untreated HIV-infected patients starting approximately 10 weeks post infection and up to seven years [286]. Similarly, they found that restriction factor expression significantly associated with systemic CD4+ T cell activation and viral load. Their study lacked acute phase measurements and also did not differentiate among CD4+ T cell subsets. The authors speculated that a general increase in restriction factor expression could be due to viral infection and depletion of cells expressing lower levels of restriction factors, but this conclusion is hard to verify without analyzing expression in specific CD4+ T cell subsets. Our post-infection data shows that memory subsets, despite expressing relatively higher levels of restriction factors than did naïve, become infected at a higher rate, so restriction factor expression levels alone are unlikely to explain differential susceptibility.

Tissue differences in gene expression

No study published to date has measured either restriction factor expression in CD4+ T cells from the intestinal mucosa or restriction factor expression during acute infection. The initial target cells of most human HIV infections are located at mucosal sites, yet potential tissue differences have been understudied due to limited accessibility in humans. Mucosal CD4+ T cells are particularly susceptible to infection and are depleted early in both acute SIV and HIV infection. One hypothesis proposed to explain this observation is the higher proportion of memory cells (and especially CCR5-expressing cells) than naïve in mucosal sites, which we also observed by flow cytometry. Another possibility is a higher proportion of "activated" cells. Although cellular activation has multiple definitions, a predicted phenotype of activated cells would be generally higher levels of transcription. The non-normalized Ct data for equal cell

equivalent inputs from peripheral blood and jejunum in the same subset did not significantly differ. For example, the mean Ct was 18.6 in peripheral blood TM, which is slightly higher expression than the mean Ct of 18.9 in jejunum TM cells across 182 expressed genes. Since we did not see evidence of overall increases in transcription in tightly defined populations of memory cells in jejunum compared to peripheral blood, the observed increase in activation between tissues may mainly reflect the differing proportions of memory subsets. It is possible that overall transcription is not the best measure of activation and that some other aspect of activation underlies tissue differences, or that cell processing altered the true *in vivo* conditions.

In light of their importance in HIV acquisition and pathogenesis, there has been significant interest in designing an HIV vaccine that prevents infection at mucosal sites (reviewed in [393]). Expression of restriction factors has correlated with reduction of viremia in IFN-α treatment *in vivo* in both humans and macaques [203,206], expression of the likely restriction factor SLFN11 correlated with elite control [285], and higher basal expression of TRIM5α correlated with resistance to repeated low dose SIV challenge [394], cumulatively suggesting a mechanism in which increased expression of restriction factors could inhibit infection. Future vaccine or adjuvant studies could measure restriction factor expression to determine if expression can be induced by vaccines, how durable induction is, and if expression of restriction factors correlates with protection.

Post-infection restriction factor expression

At the peak of acute SIV infection, expression of many restriction factors was increased compared to uninfected animals in both peripheral blood and jejunum, suggesting systemic modulation of expression in multiple compartments. While known interferon-stimulated genes increased the most during infection, other non-ISGs also exhibited altered expression; therefore,

it is likely cytokines/chemokines in addition to interferon play a role *in vivo*. The largely unknown mechanisms of interferon-independent control of restriction factor expression could have important implications, for example in vaccine/adjuvant design.

Interestingly, the fact that transitional memory (TM) CCR5+ CD4+ T cells from jejunum expressed relatively lower levels of restriction factors than the same cells from peripheral blood also held true at 10 days post infection (data not shown). Cells from jejunum were more highly infected than peripheral blood, yet the jejunum TM cells still expressed lower levels of restriction factors relative to TM cells in peripheral blood. This suggests that the level of SIV infection of specific CD4+ T cell subsets is not the major driver of changes in expression of restriction factors, and that other cell types such as dendritic cells may be the source of the cytokine cascade known to occur in acute infection [199] that could drive increased expression of restriction factors.

4.C. Single cell expression profiling

Single cell expression technologies

A number of factors led to the selection of qRT-PCR for single cell expression analysis used for these studies. qRT-PCR allows for high-throughput, highly quantitative, and precise gene expression measurements with a wide dynamic range of 6-8 orders of magnitude, significantly wider than microarray based systems [325]. In addition, the specificity of the primer/probe real-time PCR assays permits discrimination of specific transcript isoforms. The reasonable per-cell cost allows for measurement of hundreds of cells, which is necessary for obtaining a sufficient number of infected cells to perform comparative analysis.

A drawback to qRT-PCR is that target genes must be selected prior to analysis and that assay throughput is limited. Microarrays are a genome-wide technology that provide a more comprehensive analysis of gene expression. At the time we began this work, microarrays from single cells were technically possible [395-397]. However, the large amount of pre-amplification necessary to generate sufficient cDNA from a single cell could lead to significant bias, and the sensitivity, specificity, and flexibility of microarrays is inferior to qRT-PCR [277]. For example, a published method for single cell microarrays consists of 30 total cycles of PCR amplification prior to array hybridization [396]. Splitting the cDNA from a single cell to 96 individual reaction chambers in a BioMark IFC also requires pre-amplification. However, to ensure a single mRNA is distributed into each chamber, a minimum of 12 cycles are required, and 18 has been shown to be reliably sufficient [363]. Gene specific pre-amplification as used in microfluidic qRT-PCR is also less likely to generate non-specific amplification as shown by the need for an exonuclease step in microarray amplification [396]. In addition, the high per-cell expense would limit throughput of cells, limiting statistical power.

More recently, single cell RNA sequencing (scRNAseq) has become a popular tool for estimating gene expression by counting reads aligning to a coding region of a gene [398,399]. Benefits to this technology include potential transcriptome-wide coverage and the ability to detect microRNAs, splice variants, and uncharacterized transcripts. The initial scRNAseq publication performed a comparison with single cell microarrays and found 75% more reads and over a thousand unknown splice junctions [398]. Despite these potential advantages, technical challenges at the time this project was started prevented use of this technology as did a number of significant inherent drawbacks. As with microarrays, significant amplification of cDNA is required that introduces bias and non-specific products, such that findings using microarray or

RNAseq generally require validation by qPCR or *in situ* hybridization staining [277]. The published and commercially available methods for reverse transcription and barcoding required for scRNAseq rely on poly-T priming [400,401], and therefore would not detect non-poly-adenylated viral sequences (for example, the abortive R-region transcripts seen in latently HIV infected cells [402]). The per-cell cost for scRNAseq is at least double that of single cell qRT-PCR, and data analysis is much more complicated. Finally, the sensitivity of qRT-PCR in the Fluidigm Biomark system has been shown to be able to distinguish a 1.25 fold difference in relative quantity [403] with a limit of detection of a single mRNA transcript [363], metrics which scRNAseq likely cannot match since it has been estimated that about 10 transcript copies per cell cannot be distinguished from noise [404].

Single cell expression of ISGs

An interesting observation in the single cell gene expression data was an enrichment of detection of type I interferon-stimulated gene expression in infected single cells compared to uninfected cells **(Figure 3.3.D)**. The median rank of type I ISGs is 23 of out 85 cellular genes based on discrete expression, including seven of the top 10 genes. For example, a cell that expresses detectable *oas1* was determined to be infected 34.6% of the time compared to a 20% average for any given gene, and just missed significance as an outlier (p=0.08). On average, all ISG genes increased expression as measured by both discrete and continuous expression in infected compared to uninfected cells. Since infected cells but not uninfected bystanders from the same tissue increased expression of ISGs (but not overall expression, **Figure3.3.B** and **C**), this suggests a cell-intrinsic mechanism of transcriptional regulation and not systemic signals that would affect both infected cells and their uninfected neighbors. Future studies will be needed to assess the mechanisms of direct viral induction of ISGs in individual infected CD4+ T cells.

Some possibilities include sensing of viral capsid by TRIM5α [124], sensing viral RT products by IFI16 [196], and sensing of viral particle assembly/budding by BST-2/Tetherin [198]. The signaling cascade that leads to interferon production through the transcription factor IRF3 can also result in expression of ISGs without prior interferon signaling [405,406].

Gene expression: cause or consequence of infection

Since qRT-PCR is a snapshot of a particular cell at a particular time, it is not possible to track changes in gene expression in a single cell over time. When comparing infected and uninfected cells, there is a chance that infection altered the expression of cellular genes, as posited above in regard to ISGs. We have evidence suggesting that PD-1 and CXCR5 expression is not altered by infection at a bulk level by comparing CD4+ T cells from uninfected animals to those from acutely SIV-infected animals. It may be possible to further characterize this at a single cell level *in vitro* by comparing expression levels in many cells before and after infection, with the caveat that the same cell cannot be analyzed twice.

4.D. Identification of a novel PD-1+ CXCR5+ CD4+ T cell population in the jejunum—remaining questions

Localization and normal biological functions

The precise localization of PD-1+ CXCR5+ CD4+ T cells within the jejunum is not currently known. Differences in genes that determine germinal center positioning (CCR7, GPR183/EBI2, and SAP) between the jejunum population and GC Tfh cells suggests that these cells are not located in germinal centers. The general lack of Peyer's patches in rhesus jejunum tissue supports this conclusion as well [151]. One possibility is that these are circulating Tfh cells like those found in peripheral blood; however, this jejunum population differs from circulating

Tfh cells that do not generally express BCL6 protein or the cell surface activation markers PD-1, CD69, and ICOS [382]. While macroscopically visible Peyer's patches are rare in jejunum since their size and density increases from duodenum to ileum [386,407], scattered small lymphoid nodules are occasionally present in histological sections from jejunum [151]. It is possible that these diffuse lymphoid aggregates harbor this jejunum Tfh-like population.

Tfh cells found in lymphoid tissue but not in germinal centers could be newly activated Tfh cells in the process of transiting to a germinal center or germinal center Tfh cells transiting to a new germinal center. Very little is known about transiting Tfh cells and what markers could define them. These cells could also be Tfh cells destined to provide help outside of germinal centers, a role that has been proposed [152] but little studied.

Since the populations of PD-1+ CXCR5+ memory CD4+ T cells in jejunum has not been previously described, the normal biological functions of this population are not known. Other than localization markers, we showed very strong similarity in gene expression profiles of PD-1+ CXCR5+ jejunal CD4+ T cells with germinal center Tfh cells from lymph nodes. Tfh cells in Peyer's patches are known to be tightly involved in IgA antibody responses [408]. IgA comprises the majority of all immunoglobulin produced in mammals, and since it can be secreted, is important for neutralizing pathogens and controlling commensal microbiota in intestinal lumen [409]. The general process of an IgA response involves induction in Peyer's patches, plasmablast recirculation and homing to intestinal mucosa, differentiation to plasma cells producing IgA, and secretion of IgA through the intestinal epithelium [409]. The majority of IgA sequences in gut lamina propria are highly mutated suggesting a germinal center origin, and Peyer's patch germinal center formation and maintenance is dependent on CD4+ T cells [408].

In addition to the highly mutated germinal center B cells, some T-dependent B cell differentiation occurs extrafollicularly (in the outer T zone of secondary lymphoid tissues) and produces short lived plasmablasts producing unmutated antibody that can be important for early protection against microbes [410]. Like germinal center Tfh cells, the T cells priming B cells extrafollicularly have been shown to require BCL6 and IL-21 expression [411]. However, unlike the jejunum Tfh cells we describe, these cells are PD-1^{low} [411].

Consequences of SIV infection

Many unanswered questions remain regarding possible consequences of the PD-1+ CXCR5+ cells becoming infected with SIV, especially since their function is not currently known. We currently do not know if these cells become depleted during SIV infection, though the observed level of direct viral infection and non-statistically significant decreases comparing uninfected to seven and 10 day post infection animals **(Supplemental Figure 3.2)** suggest they do. Additionally, a study failed to find PD-1^{high} CD4+ T cells in the jejunum of chronically infected animals [350], but without characterizing uninfected animals as a comparison, they could not conclude whether this was true depletion or lack of PD-1^{high} cells in normal conditions. Acute SIV infection depleted Tfh cells from the spleen [272] and lymph nodes [412]; however, other studies have observed an accumulation of Tfh cells in some animals during chronic SIV infection despite high levels of infection [412,413], and depletion again in terminal stages of disease [412]. Studying the dynamics of this population over the course of infection will thus be important.

A critical remaining question is if these cells become latently infected. The identity of the latent reservoir has been under intense investigation since knowing the characteristics of cells that are a source of virus could lead to targeted strategies designed to eliminate them. There are many recent studies implicating Tfh cells in lymph nodes as latently infected in SIV and HIV

infections [271,391,414], but since the population described here does not reside in a lymph node protected from CTL responses, it is unclear if they persist while infected.

Implications for pathogenesis

The extremely high level of infection of the PD-1+ CXCR5+ cells from jejunum suggest that they are highly susceptible to infection. They could be initial target cells in intravenous infection or, if they are also located in rectal/vaginal mucosa, a target in sexual transmission. Their location in intestinal mucosa places them in close proximity to other activated CD4+ T cells to which infection could be spread, assuming their high level of viral RNA production results in similarly high levels of virion production.

Infection that results in depletion could impair their normal functions. For example, if these cells are newly activated Tfh cells destined to transit to a lymph node to induce germinal center formation, this process could be disrupted. IgA responses in both humans and chimpanzees with HIV as well as macaques with SIV are weak or absent despite excellent IgG responses [415-418] while other mucosal infections induce preferential IgA responses [419]. Depletion of mucosal Tfh cells could explain this dichotomy. Additionally, a major source of pathogenesis in HIV appears to be due to microbial translocation (reviewed in [181]), and loss of IgA responses could contribute to the loss of barrier function seen early in infection.

Implications for prevention/treatment

Better understanding of the functions of these PD-1+ CXCR5+ cells could impact prevention and treatment of HIV. Preventing their infection though a vaccine or drug treatment could result in better antibody responses, importantly perhaps highly hypermutated broadly neutralizing antibodies. It would be interesting to determine if antiretroviral treatment can prevent their infection since evidence of mucosal replication of HIV is sometimes still seen even

in patients on treatment [420,421]. Finally, if these cells are latently infected, understanding what approaches can eliminate them could impact HIV cure research.

4.E. Future Directions

Further characterization of the PD-1+ CXCR5+ population from jejunum is currently underway. To determine if this population produces IL-21 protein, a functional mark of Tfh cells, we are currently optimizing an intracellular cytokine staining panel for IL-21, IL-2, and IFN-γ. Four uninfected jejunum samples have been obtained for testing. We anticipate they will produce IL-21 when stimulated since they produce IL-21 mRNA **(Figure 3.5.B)**.

Another critical aspect of the study involves identification of the localization of PD-1+ CXCR5+ CD4+ T cells in the gut. Jejunum and lymph node formalin-fixed and paraffin-embedded tissue samples from uninfected and SIV-infected animals that have been sent to the lab of Ashley Haase (University of Minnesota) for immunohistochemistry analysis. We plan to analyze the tissue architecture with respect to the location of this cell population and use *in situ* hybridization to identify virus-infected cells.

To verify the susceptibility of the PD-1+ CXCR5+ population to SIV infection and to more rigorously test if infection changes expression of PD-1 or CXCR5, we plan to perform an *in vitro* infection with GFP-tagged SIV. We will assess gene expression before and after infection, and measure the extent of infection in defined PD-1 and CXCR5 subpopulations.

Finally, if it is possible to find a source of samples, determining if there is an equivalent population in humans would be important to demonstrate the relevance of our findings.

References

1 **Centers for Disease, C.** Pneumocystis pneumonia--Los Angeles. *MMWR Morb Mortal Wkly Rep* **30**, 250-252 (1981).

2 **Barre-Sinoussi, F. *et al.*** Isolation of a T-lymphotrophic retrovirus from a patient at risk for acquired immunodeficiency syndrome (AIDS). *Science* **220**, 868-871 (1983).

3 **Gallo, R. C. *et al.*** Frequent Detection and Isolation of Cytopathic Retroviruses (Htlv-Iii) from Patients with Aids and at Risk for Aids. *Science* **224**, 500-503, doi:DOI 10.1126/science.6200936 (1984).

4 **Popovic, M., Sarangadharan, M. G., Read, E. & Gallo, R. C.** Detection, Isolation and continuing production of cytopathic retroviruses(HTLV III) from patients with AIDS and preAIDS. *Science* **224**, 497-500 (1984).

5 UNAIDS. HIV estimates with uncertainty bounds 1990-2015. (2016).

6 UNAIDS. Global AIDS Update 2016., 1-16 (Joint United Nations Programme on HIV/AIDS, 2016).

7 **Hall, H. I. *et al.*** Prevalence of Diagnosed and Undiagnosed HIV Infection--United States, 2008-2012. *MMWR Morb Mortal Wkly Rep* **64**, 657-662 (2015).

8 CDC. HIV Surveillance Report, 2014. (Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services, Atlanta, Georgia, 2015).

9 **Arts, E. J. & Hazuda, D. J.** HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med* **2**, a007161, doi:10.1101/cshperspect.a007161 (2012).

10 **Autran, B. *et al.*** Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. *Science* **277**, 112-116 (1997).

11 **May, M. T. *et al.*** Impact on life expectancy of HIV-1 positive individuals of CD4+ cell count and viral load response to antiretroviral therapy. *AIDS* **28**, 1193-1202, doi:10.1097/QAD.0000000000000243 (2014).

12 **Schacker, T., Collier, A. C., Hughes, J., Shea, T. & Corey, L.** Clinical and epidemiologic features of primary HIV infection. *Ann Intern Med* **125**, 257-264 (1996).

13 **Lackner, A. A., Lederman, M. M. & Rodriguez, B.** HIV pathogenesis: the host. *Cold Spring Harb Perspect Med* **2**, a007005, doi:10.1101/cshperspect.a007005 (2012).

14 **Rines, D. R. *et al.*** High-content screening of functional genomic libraries. *Methods Enzymol* **414**, 530-565, doi:10.1016/S0076-6879(06)14028-8 (2006).

15 **Zhou, H. *et al.*** Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4**, 495-504, doi:10.1016/j.chom.2008.10.004 (2008).

16 **Konig, R. *et al.*** Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135**, 49-60, doi:10.1016/j.cell.2008.07.032 (2008).

17 **Brass, A. L. *et al.*** Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**, 921-926, doi:10.1126/science.1152725 (2008).

18 **Yeung, M. L., Houzet, L., Yedavalli, V. S. & Jeang, K. T.** A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J Biol Chem* **284**, 19463-19473, doi:10.1074/jbc.M109.010033 (2009).

19 **Bushman, F. D. *et al.*** Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog* **5**, e1000437, doi:10.1371/journal.ppat.1000437 (2009).

20 **Goff, S. P.** Knockdown screens to knockout HIV-1. *Cell* **135**, 417-420, doi:10.1016/j.cell.2008.10.007 (2008).

21 **Zhu, J. *et al.*** Comprehensive identification of host modulators of HIV-1 replication using multiple orthologous RNAi reagents. *Cell Rep* **9**, 752-766, doi:10.1016/j.celrep.2014.09.031 (2014).

22 **Saphire, A. C., Bobardt, M. D., Zhang, Z., David, G. & Gallay, P. A.** Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J Virol* **75**, 9187-9200, doi:10.1128/JVI.75.19.9187-9200.2001 (2001).

23 **Geijtenbeek, T. B. *et al.*** DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587-597 (2000).

24 **Arthos, J. *et al.*** HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. *Nat Immunol* **9**, 301-309, doi:10.1038/ni1566 (2008).

25 **Cicala, C. *et al.*** The integrin alpha4beta7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1. *Proc Natl Acad Sci U S A* **106**, 20877-20882, doi:10.1073/pnas.0911796106 (2009).

26 **Maddon, P. J. *et al.*** The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333-348 (1986).

27 **McDougal, J. S. *et al.*** Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**, 382-385 (1986).

28 **Hartley, O., Klasse, P. J., Sattentau, Q. J. & Moore, J. P.** V3: HIV's switch-hitter. *AIDS Res Hum Retroviruses* **21**, 171-189, doi:10.1089/aid.2005.21.171 (2005).

29 **Berger, E. A. *et al.*** A new classification for HIV-1. *Nature* **391**, 240, doi:10.1038/34571 (1998).

30 **Chen, Z. W., Gettie, A., Ho, D. D. & Marx, P. A.** Primary SIVsm isolates use the CCR5 coreceptor from sooty mangabeys naturally infected in west Africa: A comparison of coreceptor usage of primary SIVsm, HIV-2, and SIVmac. *Virology* **246**, 113-124, doi:DOI 10.1006/viro.1998.9174 (1998).

31 **Farzan, M. *et al.*** Two orphan seven-transmembrane segment receptors which are expressed in CD4-positive cells support simian immunodeficiency virus infection. *Journal of Experimental Medicine* **186**, 405-411, doi:DOI 10.1084/jem.186.3.405 (1997).

32 **Wilen, C. B., Tilton, J. C. & Doms, R. W.** HIV: cell binding and entry. *Cold Spring Harb Perspect Med* **2**, doi:10.1101/cshperspect.a006866 (2012).

33 **Splettstoesser, T.** *Schematic description of the HIV replication cycle.*, <https://commons.wikimedia.org/wiki/File:HIV-replication-cycle.svg> (2014).

34 **Hu, W. S. & Hughes, S. H.** HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**, doi:10.1101/cshperspect.a006882 (2012).

35 **Wakefield, J. K., Wolf, A. G. & Morrow, C. D.** Human immunodeficiency virus type 1 can use different tRNAs as primers for reverse transcription but selectively maintains a primer binding site complementary to tRNA(3Lys). *J Virol* **69**, 6021-6029 (1995).

36 **Charneau, P., Alizon, M. & Clavel, F.** A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J Virol* **66**, 2814-2820 (1992).

37 **Whitcomb, J. M., Kumar, R. & Hughes, S. H.** Sequence of the circle junction of human immunodeficiency virus type 1: implications for reverse transcription and integration. *J Virol* **64**, 4903-4906 (1990).

38 **Iordanskiy, S. & Bukrinsky, M.** Reverse transcription complex: the key player of the early phase of HIV replication. *Future Virol* **2**, 49-64, doi:10.2217/17460794.2.1.49 (2007).

39 **Arhel, N.** Revisiting HIV-1 uncoating. *Retrovirology* **7**, 96, doi:10.1186/1742-4690-7-96 (2010).

40 **Miller, M. D., Farnet, C. M. & Bushman, F. D.** Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* **71**, 5382-5390 (1997).

41 **Hulme, A. E., Kelley, Z., Foley, D. & Hope, T. J.** Complementary Assays Reveal a Low Level of CA Associated with Viral Complexes in the Nuclei of HIV-1-Infected Cells. *J Virol* **89**, 5350-5361, doi:10.1128/JVI.00476-15 (2015).

42 **Luban, J., Bossolt, K. L., Franke, E. K., Kalpana, G. V. & Goff, S. P.** Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* **73**, 1067-1078 (1993).

43 **Franke, E. K., Yuan, H. E. & Luban, J.** Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **372**, 359-362, doi:10.1038/372359a0 (1994).

44 **Gamble, T. R. *et al.*** Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285-1294 (1996).

45 **Thali, M. *et al.*** Functional association of cyclophilin A with HIV-1 virions. *Nature* **372**, 363-365, doi:10.1038/372363a0 (1994).

46 **Hatziioannou, T., Perez-Caballero, D., Cowan, S. & Bieniasz, P. D.** Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells. *J Virol* **79**, 176-183, doi:10.1128/JVI.79.1.176-183.2005 (2005).

47 **Campbell, E. M. & Hope, T. J.** HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat Rev Microbiol* **13**, 471-483, doi:10.1038/nrmicro3503 (2015).

48 **Farnet, C. M. & Bushman, F. D.** HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* **88**, 483-492 (1997).

49 **Llano, M. *et al.*** LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J Virol* **78**, 9524-9537, doi:10.1128/JVI.78.17.9524-9537.2004 (2004).

50 **Thys, W. *et al.*** Interplay between HIV entry and transportin-SR2 dependency. *Retrovirology* **8**, 7, doi:10.1186/1742-4690-8-7 (2011).

51 **De Iaco, A. *et al.*** TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology* **10**, 20, doi:10.1186/1742-4690-10-20 (2013).

52 **Lee, K. *et al.*** HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. *J Virol* **86**, 3851-3860, doi:10.1128/JVI.06607-11 (2012).

53 **Zhang, R., Mehla, R. & Chauhan, A.** Perturbation of host nuclear membrane component RanBP2 impairs the nuclear import of human immunodeficiency virus -1 preintegration complex (DNA). *PLoS One* **5**, e15620, doi:10.1371/journal.pone.0015620 (2010).

54 **Ocwieja, K. E. *et al.*** HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog* **7**, e1001313, doi:10.1371/journal.ppat.1001313 (2011).

55 **Matreyek, K. A. & Engelman, A.** The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. *J Virol* **85**, 7818-7827, doi:10.1128/JVI.00325-11 (2011).

56 **Craigie, R. & Bushman, F. D.** HIV DNA integration. *Cold Spring Harb Perspect Med* **2**, a006890, doi:10.1101/cshperspect.a006890 (2012).

57 **Suzuki, Y. & Craigie, R.** The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol* **5**, 187-196, doi:10.1038/nrmicro1579 (2007).

58 **Wu, X., Li, Y., Crise, B. & Burgess, S. M.** Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749-1751, doi:10.1126/science.1083413 (2003).

59 **Mitchell, R. S. *et al.*** Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* **2**, E234, doi:10.1371/journal.pbio.0020234 (2004).

60 **Maertens, G. *et al.*** LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem* **278**, 33528-33539, doi:10.1074/jbc.M303594200 (2003).

61 **Cherepanov, P. *et al.*** HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* **278**, 372-381, doi:10.1074/jbc.M209278200 (2003).

62 **Christ, F. & Debyser, Z.** The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy. *Virology* **435**, 102-109, doi:10.1016/j.virol.2012.09.033 (2013).

63 **Sodroski, J. *et al.*** Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. *Science* **227**, 171-173 (1985).

64 **Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W.** Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. *Science* **229**, 74-77 (1985).

65 **Herrmann, C. H., Gold, M. O. & Rice, A. P.** Viral transactivators specifically target distinct cellular protein kinases that phosphorylate the RNA polymerase II C-terminal domain. *Nucleic Acids Res* **24**, 501-508 (1996).

66 **Herrmann, C. H. & Rice, A. P.** Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J Virol* **69**, 1612-1620 (1995).

67 **Mancebo, H. S. *et al.*** P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev* **11**, 2633-2644 (1997).

68 **Fujinaga, K. *et al.*** Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol* **24**, 787-795 (2004).

69 **Bourgeois, C. F., Kim, Y. K., Churcher, M. J., West, M. J. & Karn, J.** Spt5 cooperates with human immunodeficiency virus type 1 Tat by preventing premature RNA release at terminator sequences. *Mol Cell Biol* **22**, 1079-1093 (2002).

70 **Kim, Y. K., Bourgeois, C. F., Isel, C., Churcher, M. J. & Karn, J.** Phosphorylation of the RNA polymerase II carboxyl-terminal domain by CDK9 is directly responsible for human immunodeficiency virus type 1 Tat-activated transcriptional elongation. *Mol Cell Biol* **22**, 4622-4637 (2002).

71 **Nabel, G. & Baltimore, D.** An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* **326**, 711-713, doi:10.1038/326711a0 (1987).

72 **Liu, J., Perkins, N. D., Schmid, R. M. & Nabel, G. J.** Specific NF-kappa B subunits act in concert with Tat to stimulate human immunodeficiency virus type 1 transcription. *J Virol* **66**, 3883-3887 (1992).

73 **Kinoshita, S., Chen, B. K., Kaneshima, H. & Nolan, G. P.** Host control of HIV-1 parasitism in T cells by the nuclear factor of activated T cells. *Cell* **95**, 595-604 (1998).

74 **Alcami, J. *et al.*** Absolute dependence on kappa B responsive elements for initiation and Tat-mediated amplification of HIV transcription in blood CD4 T lymphocytes. *EMBO J* **14**, 1552-1560 (1995).

75 **Ruiz, A. *et al.*** Characterization of the influence of mediator complex in HIV-1 transcription. *J Biol Chem* **289**, 27665-27676, doi:10.1074/jbc.M114.570341 (2014).

76 **Ballana, E. *et al.*** ZNRD1 (zinc ribbon domain-containing 1) is a host cellular factor that influences HIV-1 replication and disease progression. *Clin Infect Dis* **50**, 1022-1032, doi:10.1086/651114 (2010).

77 **Sodroski, J. *et al.*** A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature* **321**, 412-417, doi:10.1038/321412a0 (1986).

78 **Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V. & Cullen, B. R.** The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254-257, doi:10.1038/338254a0 (1989).

79 **Pollard, V. W. & Malim, M. H.** The HIV-1 Rev protein. *Annu Rev Microbiol* **52**, 491-532, doi:10.1146/annurev.micro.52.1.491 (1998).

80 **Karn, J. & Stoltzfus, C. M.** Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med* **2**, a006916, doi:10.1101/cshperspect.a006916 (2012).

81 **Liu, J., Henao-Mejia, J., Liu, H., Zhao, Y. & He, J. J.** Translational regulation of HIV-1 replication by HIV-1 Rev cellular cofactors Sam68, eIF5A, hRIP, and DDX3. *J Neuroimmune Pharmacol* **6**, 308-321, doi:10.1007/s11481-011-9265-8 (2011).

82 **Camus, G. *et al.*** The clathrin adaptor complex AP-1 binds HIV-1 and MLV Gag and facilitates their budding. *Mol Biol Cell* **18**, 3193-3203, doi:10.1091/mbc.E06-12-1147 (2007).

83 **Batonick, M. *et al.*** Interaction of HIV-1 Gag with the clathrin-associated adaptor AP-2. *Virology* **342**, 190-200, doi:10.1016/j.virol.2005.08.001 (2005).

84 **Dong, X. *et al.*** AP-3 directs the intracellular trafficking of HIV-1 Gag and plays a key role in particle assembly. *Cell* **120**, 663-674, doi:10.1016/j.cell.2004.12.023 (2005).

85 **Tang, Y. *et al.*** Cellular motor protein KIF-4 associates with retroviral Gag. *J Virol* **73**, 10508-10513 (1999).

86 **Dooher, J. E., Schneider, B. L., Reed, J. C. & Lingappa, J. R.** Host ABCE1 is at plasma membrane HIV assembly sites and its dissociation from Gag is linked to subsequent events of virus production. *Traffic* **8**, 195-211, doi:10.1111/j.1600-0854.2006.00524.x (2007).

87 **Johnson, S. F. & Telesnitsky, A.** Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathog* **6**, e1001007, doi:10.1371/journal.ppat.1001007 (2010).

88 **Morita, E. & Sundquist, W. I.** Retrovirus budding. *Annu Rev Cell Dev Biol* **20**, 395-425, doi:10.1146/annurev.cellbio.20.010403.102350 (2004).

89 **Garrus, J. E. *et al.*** Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**, 55-65 (2001).

90 **Strack, B., Calistri, A., Craig, S., Popova, E. & Gottlinger, H. G.** AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **114**, 689-699 (2003).

91 **Jouvenet, N., Zhadina, M., Bieniasz, P. D. & Simon, S. M.** Dynamics of ESCRT protein recruitment during retroviral assembly. *Nat Cell Biol* **13**, 394-401, doi:10.1038/ncb2207 (2011).

92 **Harris, R. S., Hultquist, J. F. & Evans, D. T.** The restriction factors of human immunodeficiency virus. *J Biol Chem* **287**, 40875-40883, doi:10.1074/jbc.R112.416925 (2012).

93 **Casazza, J. P. *et al.*** Autocrine production of beta-chemokines protects CMV-Specific CD4 T cells from HIV infection. *PLoS Pathog* **5**, e1000646, doi:10.1371/journal.ppat.1000646 (2009).

94 **Lu, J. *et al.*** The IFITM proteins inhibit HIV-1 infection. *J Virol* **85**, 2126-2137, doi:10.1128/JVI.01531-10 (2011).

95 **Schoggins, J. W. *et al.*** A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481-485, doi:10.1038/nature09907 (2011).

96 **Li, K. *et al.*** IFITM proteins restrict viral membrane hemifusion. *PLoS Pathog* **9**, e1003124, doi:10.1371/journal.ppat.1003124 (2013).

97 **Yu, J. *et al.*** IFITM Proteins Restrict HIV-1 Infection by Antagonizing the Envelope Glycoprotein. *Cell Rep* **13**, 145-156, doi:10.1016/j.celrep.2015.08.055 (2015).

98 **Zimin, A. V. *et al.*** A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biol Direct* **9**, 20, doi:10.1186/1745-6150-9-20 (2014).

99 **Gibbs, R. A. *et al.*** Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222-234, doi:10.1126/science.1139247 (2007).

100 **Wilkins, J., Zheng, Y. M., Yu, J., Liang, C. & Liu, S. L.** Nonhuman Primate IFITM Proteins Are Potent Inhibitors of HIV and SIV. *PLoS One* **11**, e0156739, doi:10.1371/journal.pone.0156739 (2016).

101 **Usami, Y., Wu, Y. & Gottlinger, H. G.** SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **526**, 218-223, doi:10.1038/nature15400 (2015).

102 **Rosa, A. *et al.*** HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **526**, 212-217, doi:10.1038/nature15399 (2015).

103 **Murrell, B., Vollbrecht, T., Guatelli, J. & Wertheim, J. O.** The Evolutionary Histories of Antiretroviral Proteins SERINC3 and SERINC5 Do Not Support an Evolutionary Arms Race in Primates. *J Virol* **90**, 8085-8089, doi:10.1128/JVI.00972-16 (2016).

104 **Laguette, N. *et al.*** SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654-657, doi:10.1038/nature10117 (2011).

105 **Hrecka, K. *et al.*** Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**, 658-661, doi:10.1038/nature10195 (2011).

106 **Diamond, T. L. *et al.*** Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem* **279**, 51545-51553, doi:10.1074/jbc.M408573200 (2004).

107 **Descours, B. *et al.*** SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4(+) T-cells. *Retrovirology* **9**, 87, doi:10.1186/1742-4690-9-87 (2012).

108 **Baldauf, H. M. *et al.*** SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nat Med* **18**, 1682-1687, doi:10.1038/nm.2964 (2012).

109 **Fregoso, O. I. *et al.*** Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the restriction factor SAMHD1. *PLoS Pathog* **9**, e1003496, doi:10.1371/journal.ppat.1003496 (2013).

110 **Bergamaschi, A. *et al.*** The human immunodeficiency virus type 2 Vpx protein usurps the CUL4A-DDB1 DCAF1 ubiquitin ligase to overcome a postentry block in macrophage infection. *J Virol* **83**, 4854-4860, doi:10.1128/JVI.00187-09 (2009).

111 **Schwefel, D. *et al.*** Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **505**, 234-238, doi:10.1038/nature12815 (2014).

112 **Ahn, J. *et al.*** HIV/simian immunodeficiency virus (SIV) accessory virulence factor Vpx loads the host cell restriction factor SAMHD1 onto the E3 ubiquitin ligase complex CRL4DCAF1. *J Biol Chem* **287**, 12550-12558, doi:10.1074/jbc.M112.340711 (2012).

113 **Pauls, E. *et al.*** Cell cycle control and HIV-1 susceptibility are linked by CDK6-dependent CDK2 phosphorylation of SAMHD1 in myeloid and lymphoid cells. *J Immunol* **193**, 1988-1997, doi:10.4049/jimmunol.1400873 (2014).

114 **Allouch, A. *et al.*** p21-mediated RNR2 repression restricts HIV-1 replication in macrophages by inhibiting dNTP biosynthesis pathway. *Proc Natl Acad Sci U S A* **110**, E3997-4006, doi:10.1073/pnas.1306719110 (2013).

115 **Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H.** Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650, doi:10.1038/nature00939 (2002).

116 **Hultquist, J. F. *et al.*** Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* **85**, 11220-11234, doi:10.1128/JVI.05238-11 (2011).

117 **Larue, R. S., Lengyel, J., Jonsson, S. R., Andresdottir, V. & Harris, R. S.** Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. *J Virol* **84**, 8193-8201, doi:10.1128/JVI.00685-10 (2010).

118 **Yu, X. *et al.*** Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-1060, doi:10.1126/science.1089591 (2003).

119 **Liu, L. *et al.*** A whole genome screen for HIV restriction factors. *Retrovirology* **8**, 94, doi:10.1186/1742-4690-8-94 (2011).

120 **Marno, K. M. *et al.*** Novel restriction factor RNA-associated early-stage anti-viral factor (REAF) inhibits human and simian immunodeficiency viruses. *Retrovirology* **11**, 3, doi:10.1186/1742-4690-11-3 (2014).

121 **Stremlau, M. *et al.*** The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).

122 **Stremlau, M. *et al.*** Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad Sci U S A* **103**, 5514-5519, doi:10.1073/pnas.0509996103 (2006).

123 **Wu, X., Anderson, J. L., Campbell, E. M., Joseph, A. M. & Hope, T. J.** Proteasome inhibitors uncouple rhesus TRIM5alpha restriction of HIV-1 reverse transcription and infection. *Proc Natl Acad Sci U S A* **103**, 7465-7470, doi:10.1073/pnas.0510483103 (2006).

124 **Pertel, T. *et al.*** TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472**, 361-365, doi:10.1038/nature09976 (2011).

125 **Kirmaier, A. *et al.*** TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. *PLoS Biol* **8**, doi:10.1371/journal.pbio.1000462 (2010).

126 **Hatziioannou, T. *et al.*** Generation of simian-tropic HIV-1 by restriction factor evasion. *Science* **314**, 95, doi:10.1126/science.1130994 (2006).

127 **Uchil, P. D., Quinlan, B. D., Chan, W. T., Luna, J. M. & Mothes, W.** TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathog* **4**, e16, doi:10.1371/journal.ppat.0040016 (2008).

128 **Carthagena, L. *et al.*** Human TRIM gene expression in response to interferons. *PLoS One* **4**, e4894, doi:10.1371/journal.pone.0004894 (2009).

129 **Kajaste-Rudnitski, A. *et al.*** TRIM22 inhibits HIV-1 transcription independently of its E3 ubiquitin ligase activity, Tat, and NF-kappaB-responsive long terminal repeat elements. *J Virol* **85**, 5183-5196, doi:10.1128/JVI.02302-10 (2011).

130 **Allouch, A. *et al.*** The TRIM family protein KAP1 inhibits HIV-1 integration. *Cell Host Microbe* **9**, 484-495, doi:10.1016/j.chom.2011.05.004 (2011).

131 **Goujon, C. *et al.*** Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559-562, doi:10.1038/nature12542 (2013).

132 **Liu, Z. *et al.*** The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe* **14**, 398-410, doi:10.1016/j.chom.2013.08.015 (2013).

133 **Kane, M. *et al.*** MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563-566, doi:10.1038/nature12653 (2013).

134 **Fribourgh, J. L. *et al.*** Structural insight into HIV-1 restriction by MxB. *Cell Host Microbe* **16**, 627-638, doi:10.1016/j.chom.2014.09.021 (2014).

135 **Li, L. *et al.*** Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J* **20**, 3272-3281, doi:10.1093/emboj/20.12.3272 (2001).

136 **Jeanson, L. *et al.*** Effect of Ku80 depletion on the preintegrative steps of HIV-1 replication in human cells. *Virology* **300**, 100-108 (2002).

137 **Kilzer, J. M. *et al.*** Roles of host cell factors in circularization of retroviral dna. *Virology* **314**, 460-467 (2003).

138 **Yoder, K. *et al.*** The DNA repair genes XPB and XPD defend cells from retroviral infection. *Proc Natl Acad Sci U S A* **103**, 4622-4627, doi:10.1073/pnas.0509828103 (2006).

139 **Liu, S. *et al.*** MCPIP1 restricts HIV infection and is rapidly degraded in activated CD4+ T cells. *Proc Natl Acad Sci U S A* **110**, 19083-19088, doi:10.1073/pnas.1316208110 (2013).

140 **Li, H. & Wang, T. T.** MCPIP1/Regnase-I Inhibits Simian Immunodeficiency Virus and Is Not Counteracted by Vpx. *J Gen Virol*, doi:10.1099/jgv.0.000482 (2016).

141 **Li, M. *et al.*** Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* **491**, 125-128, doi:10.1038/nature11433 (2012).

142 **van Weringh, A. *et al.*** HIV-1 modulates the tRNA pool to improve translation efficiency. *Mol Biol Evol* **28**, 1827-1834, doi:10.1093/molbev/msr005 (2011).

143 **Van Damme, N. *et al.*** The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe* **3**, 245-252, doi:10.1016/j.chom.2008.03.001 (2008).

144 **Neil, S. J., Zang, T. & Bieniasz, P. D.** Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425-430, doi:10.1038/nature06553 (2008).

145 **Perez-Caballero, D. *et al.*** Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* **139**, 499-511, doi:10.1016/j.cell.2009.08.039 (2009).

146 **Mitchell, R. S. *et al.*** Vpu antagonizes BST-2-mediated restriction of HIV-1 release via beta-TrCP and endo-lysosomal trafficking. *PLoS Pathog* **5**, e1000450, doi:10.1371/journal.ppat.1000450 (2009).

147 **Jia, B. *et al.*** Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. *PLoS Pathog* **5**, e1000429, doi:10.1371/journal.ppat.1000429 (2009).

148 **Zhang, F. *et al.*** SIV Nef proteins recruit the AP-2 complex to antagonize Tetherin and facilitate virion release. *PLoS Pathog* **7**, e1002039, doi:10.1371/journal.ppat.1002039 (2011).

149 **Le Tortorec, A. & Neil, S. J.** Antagonism to and intracellular sequestration of human tetherin by the human immunodeficiency virus type 2 envelope glycoprotein. *J Virol* **83**, 11966-11978, doi:10.1128/JVI.01515-09 (2009).

150 **Rogers, J. *et al.*** An initial genetic linkage map of the rhesus macaque (Macaca mulatta) genome using human microsatellite loci. *Genomics* **87**, 30-38, doi:10.1016/j.ygeno.2005.10.004 (2006).

151 **Veazey, R. S. *et al.*** Characterization of gut-associated lymphoid tissue (GALT) of normal rhesus macaques. *Clin Immunol Immunopathol* **82**, 230-242 (1997).

152 **Crotty, S.** T follicular helper cell differentiation, function, and roles in disease. *Immunity* **41**, 529-542, doi:10.1016/j.immuni.2014.10.004 (2014).

153 **Locci, M. *et al.*** Activin A programs the differentiation of human TFH cells. *Nat Immunol* **17**, 976-984, doi:10.1038/ni.3494 (2016).

154 **Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M.** AIDS - AIDS as a zoonosis: Scientific and public health implications. *Science* **287**, 607-614, doi:DOI 10.1126/science.287.5453.607 (2000).

155 **Marx, P. A. & Chen, Z.** The function of simian chemokine receptors in the replication of SIV. *Semin Immunol* **10**, 215-223, doi:10.1006/smim.1998.0135 (1998).

156 **Crise, B. *et al.*** Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J Virol* **79**, 12199-12204, doi:10.1128/JVI.79.19.12199-12204.2005 (2005).

157 **Keele, B. F. *et al.*** Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523-526, doi:10.1126/science.1126531 (2006).

158 **Sharp, P. M. & Hahn, B. H.** Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* **1**, a006841, doi:10.1101/cshperspect.a006841 (2011).

159 **Apetrei, C. *et al.*** Molecular epidemiology of simian immunodeficiency virus SIVsm in U.S. primate centers unravels the origin of SIVmac and SIVstm. *J Virol* **79**, 8991-9005, doi:10.1128/JVI.79.14.8991-9005.2005 (2005).

160 **Paiardini, M., Pandrea, I., Apetrei, C. & Silvestri, G.** Lessons learned from the natural hosts of HIV-related viruses. *Annu Rev Med* **60**, 485-495, doi:10.1146/annurev.med.60.041807.123753 (2009).

161 **Daniel, M. D. *et al.*** Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* **228**, 1201-1204 (1985).

162 **Letvin, N. L. *et al.*** Induction of AIDS-like diseases in macaque monkeys with T-cell tropic retrovirus STLV-III. *Science* **230**, 71-73 (1985).

163 **Veazey, R. S. *et al.*** The gastrointestinal tract as a major site of CD4 T lymphocyte depletion and viral replication in SIV infection. *Science* **280**, 427-431 (1998).

164 **Van Rompay, K. K.** The use of nonhuman primate models of HIV infection for the evaluation of antiviral strategies. *AIDS Res Hum Retroviruses* **28**, 16-35, doi:10.1089/AID.2011.0234 (2012).

165 **Mattapallil, J. J. *et al.*** Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature* **434**, 1093-1097, doi:10.1038/nature03501 (2005).

166 **Brenchley, J. M. *et al.*** CD4(+) T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *Journal of Experimental Medicine* **200**, 749-759, doi:10.1084/jem.20040874 (2004).

167 **Guadalupe, M. *et al.*** Viral suppression and immune restoration in the gastrointestinal mucosa of human immunodeficiency virus type 1-infected patients initiating therapy during primary or chronic infection. *J Virol* **80**, 8236-8247, doi:10.1128/JVI.00120-06 (2006).

168 **Deleage, C. *et al.*** Impact of early cART in the gut during acute HIV infection. *JCI Insight* **1**, doi:10.1172/jci.insight.87065 (2016).

169 **Jin, X. *et al.*** Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. *J Exp Med* **189**, 991-998 (1999).

170 **Regoes, R. R., Longini, I. M., Feinberg, M. B. & Staprans, S. I.** Preclinical assessment of HIV vaccines and microbicides by repeated low-dose virus challenges. *PLoS Med* **2**, e249, doi:10.1371/journal.pmed.0020249 (2005).

171 **Ling, B. *et al.*** SIV(mac) pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans. *AIDS* **16**, 1489-1496 (2002).

172 **Sigve**. *Vectorized version of graph showing HIV viral load in relation to CD4+ lymphocyte count*, <https://commons.wikimedia.org/wiki/File:Hiv-timecourse_copy.svg> (2011).

173 **Keele, B. F. *et al.*** Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* **105**, 7552-7557, doi:10.1073/pnas.0802203105 (2008).

174 **Shaw, G. M. & Hunter, E.** HIV transmission. *Cold Spring Harb Perspect Med* **2**, doi:10.1101/cshperspect.a006965 (2012).

175 **Cavarelli, M. & Scarlatti, G.** HIV-1 infection: the role of the gastrointestinal tract. *Am J Reprod Immunol* **71**, 537-542, doi:10.1111/aji.12245 (2014).

176 **Zhang, Z. Q. *et al.*** Roles of substrate availability and infection of resting and activated CD4+ T cells in transmission and acute simian immunodeficiency virus infection. *Proc Natl Acad Sci U S A* **101**, 5640-5645, doi:10.1073/pnas.0308425101 (2004).

177 **Miller, C. J. *et al.*** Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. *J Virol* **79**, 9217-9227, doi:10.1128/JVI.79.14.9217-9227.2005 (2005).

178 **Cameron, P. U. *et al.*** Dendritic cells exposed to human immunodeficiency virus type-1 transmit a vigorous cytopathic infection to CD4+ T cells. *Science* **257**, 383-387 (1992).

179 **McDonald, D. *et al.*** Recruitment of HIV and its receptors to dendritic cell-T cell junctions. *Science* **300**, 1295-1297, doi:10.1126/science.1084238 (2003).

180 **Li, Q. *et al.*** Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells. *Nature* **434**, 1148-1152, doi:10.1038/nature03513 (2005).

181 **Tincati, C., Douek, D. C. & Marchetti, G.** Gut barrier structure, mucosal immunity and intestinal microbiota in the pathogenesis and treatment of HIV infection. *AIDS Res Ther* **13**, 19, doi:10.1186/s12981-016-0103-1 (2016).

182 **Blaus, B.** <Blausen.com staff. "Blausen gallery 2014". Wikiversity Journal of Medicine. DOI:10.15347/wjm/2014.010. ISSN 20018762. - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=28909224> (2014).

183 **Klatt, N. R. *et al.*** CD8+ lymphocytes control viral replication in SIVmac239-infected rhesus macaques without decreasing the lifespan of productively infected cells. *PLoS Pathog* **6**, e1000747, doi:10.1371/journal.ppat.1000747 (2010).

184 **Swanstrom, R. & Coffin, J.** HIV-1 pathogenesis: the virus. *Cold Spring Harb Perspect Med* **2**, a007443, doi:10.1101/cshperspect.a007443 (2012).

185 **Ho, D. D. *et al.*** Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-126, doi:10.1038/373123a0 (1995).

186 **Laurent-Crawford, A. G. *et al.*** Membrane expression of HIV envelope glycoproteins triggers apoptosis in CD4 cells. *AIDS Res Hum Retroviruses* **9**, 761-773, doi:10.1089/aid.1993.9.761 (1993).

187 **Nardacci, R. *et al.*** Syncytial apoptosis signaling network induced by the HIV-1 envelope glycoprotein complex: an overview. *Cell Death Dis* **6**, e1846, doi:10.1038/cddis.2015.204 (2015).

188 **Leonard, R. *et al.*** Cytopathic effect of human immunodeficiency virus in T4 cells is linked to the last stage of virus infection. *Proc Natl Acad Sci U S A* **85**, 3570-3574 (1988).

189 **Pauza, C. D., Galindo, J. E. & Richman, D. D.** Reinfection results in accumulation of unintegrated viral DNA in cytopathic and persistent human immunodeficiency virus type 1 infection of CEM cells. *J Exp Med* **172**, 1035-1042 (1990).

190 **Sainski, A. M. *et al.*** The HIV-1-specific protein Casp8p41 induces death of infected cells through Bax/Bak. *J Virol* **85**, 7965-7975, doi:10.1128/JVI.02515-10 (2011).

191 **Varbanov, M., Espert, L. & Biard-Piechaczyk, M.** Mechanisms of CD4 T-cell depletion triggered by HIV-1 viral proteins. *AIDS Rev* **8**, 221-236 (2006).

192 **Finkel, T. H. *et al.*** Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV- infected lymph nodes. *Nature Med.* **1**, 129-134 (1995).

193 **Doitsh, G. *et al.*** Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. *Cell* **143**, 789-801, doi:10.1016/j.cell.2010.11.001 (2010).

194 **Lehner, T., Wang, Y., Whittall, T. & Seidl, T.** Innate immunity and HIV-1 infection. *Adv Dent Res* **23**, 19-22, doi:10.1177/0022034511399081 (2011).

195 **Gao, D. *et al.*** Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341**, 903-906, doi:10.1126/science.1240933 (2013).

196 **Monroe, K. M. *et al.*** IFI16 DNA sensor is required for death of lymphoid CD4 T cells abortively infected with HIV. *Science* **343**, 428-432, doi:10.1126/science.1243640 (2014).

197 **Beignon, A. S. *et al.*** Endocytosis of HIV-1 activates plasmacytoid dendritic cells via Toll-like receptor-viral RNA interactions. *J Clin Invest* **115**, 3265-3275, doi:10.1172/JCI26032 (2005).

198 **Galao, R. P., Le Tortorec, A., Pickering, S., Kueck, T. & Neil, S. J.** Innate sensing of HIV-1 assembly by Tetherin induces NFkappaB-dependent proinflammatory responses. *Cell Host Microbe* **12**, 633-644, doi:10.1016/j.chom.2012.10.007 (2012).

199 **Stacey, A. R. *et al.*** Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *J Virol* **83**, 3719-3733, doi:10.1128/JVI.01844-08 (2009).

200 **Lehner, T. *et al.*** Protective mucosal immunity elicited by targeted iliac lymph node immunization with a subunit SIV envelope and core vaccine in macaques. *Nat Med* **2**, 767-775 (1996).

201 **Cocchi, R. *et al.*** Identification of RANTES, MIP-1alpha and MIP-1beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* **270**, 1811-1815 (1995).

202 **Ho, D. D. *et al.*** Recombinant human interferon alfa-A suppresses HTLV-III replication in vitro. *Lancet* **1**, 602-604 (1985).

203 **Pillai, S. K. *et al.*** Role of retroviral restriction factors in the interferon-alpha-mediated suppression of HIV-1 in vivo. *Proc Natl Acad Sci U S A* **109**, 3035-3040, doi:10.1073/pnas.1111573109 (2012).

204 **Haase, A. T.** Perils at mucosal front lines for HIV and SIV and their hosts. *Nat Rev Immunol* **5**, 783-792, doi:10.1038/nri1705 (2005).

205 **Li, Q. *et al.*** Glycerol monolaurate prevents mucosal SIV transmission. *Nature* **458**, 1034-1038, doi:10.1038/nature07831 (2009).

206 **Sandler, N. G. *et al.*** Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression. *Nature* **511**, 601-605, doi:10.1038/nature13554 (2014).

207 **Martin, M. P. *et al.*** Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet* **31**, 429-434, doi:10.1038/ng934 (2002).

208 **Pelak, K. *et al.*** Copy number variation of KIR genes influences HIV-1 control. *PLoS Biol* **9**, e1001208, doi:10.1371/journal.pbio.1001208 (2011).

209 **DiNapoli, S. R., Hirsch, V. M. & Brenchley, J. M.** Macrophages in Progressive Human Immunodeficiency Virus/Simian Immunodeficiency Virus Infections. *J Virol* **90**, 7596-7606, doi:10.1128/JVI.00672-16 (2016).

210 **Giorgi, J. V. *et al.*** Predictive value of immunologic and virologic markers after long or short duration of HIV-1 infection. *J Acquir Immune Defic Syndr* **29**, 346-355 (2002).

211 **Reynolds, M. R. *et al.*** CD8+ T-lymphocyte response to major immunodominant epitopes after vaginal exposure to simian immunodeficiency virus: too late and too little. *J Virol* **79**, 9228-9235, doi:10.1128/JVI.79.14.9228-9235.2005 (2005).

212 **Phillips, R. E. *et al.*** Human-Immunodeficiency-Virus Genetic-Variation That Can Escape Cytotoxic T-Cell Recognition. *Nature* **354**, 453-459, doi:DOI 10.1038/354453a0 (1991).

213 **Borrow, P., Lewicki, H., Hahn, B. H., Shaw, G. M. & Oldstone, M. B.** Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J Virol* **68**, 6103-6110 (1994).

214 **Koup, R. A. *et al.*** Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol* **68**, 4650-4655 (1994).

215 **Kaslow, R. A. *et al.*** Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* **2**, 405-411 (1996).

216 **Gao, X. *et al.*** Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* **344**, 1668-1675, doi:10.1056/NEJM200105313442203 (2001).

217 **Schmitz, J. E. *et al.*** Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science* **283**, 857-860 (1999).

218 **Goonetilleke, N. *et al.*** The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J Exp Med* **206**, 1253-1272, doi:10.1084/jem.20090365 (2009).

219 **Veiga-Fernandes, H., Walter, U., Bourgeois, C., McLean, A. & Rocha, B.** Response of naive and memory CD8+ T cells to antigen stimulation in vivo. *Nat Immunol* **1**, 47-53, doi:10.1038/76907 (2000).

220 **Wong, J. K. *et al.*** In vivo CD8+ T-cell suppression of siv viremia is not mediated by CTL clearance of productively infected cells. *PLoS Pathog* **6**, e1000748, doi:10.1371/journal.ppat.1000748 (2010).

221 **Altfeld, M. *et al.*** Expansion of pre-existing, lymph node-localized CD8+ T cells during supervised treatment interruptions in chronic HIV-1 infection. *J Clin Invest* **109**, 837-843, doi:10.1172/JCI14789 (2002).

222 **Walker, B. & McMichael, A.** The T-cell response to HIV. *Cold Spring Harb Perspect Med* **2**, doi:10.1101/cshperspect.a007054 (2012).

223 **Streeck, H. *et al.*** Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells. *J Virol* **83**, 7641-7648, doi:10.1128/JVI.00182-09 (2009).

224 **Addo, M. M. *et al.*** Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J Virol* **77**, 2081-2092 (2003).

225 **Kiepiela, P. *et al.*** CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med* **13**, 46-53, doi:10.1038/nm1520 (2007).

226 **Betts, M. R. & Harari, A.** Phenotype and function of protective T cell immune responses in HIV. *Curr Opin HIV AIDS* **3**, 349-355, doi:10.1097/COH.0b013e3282fbaa81 (2008).

227 **Fellay, J. *et al.*** A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**, 944-947, doi:10.1126/science.1143767 (2007).

228 **Ramduth, D. *et al.*** Differential immunogenicity of HIV-1 clade C proteins in eliciting CD8+ and CD4+ cell responses. *J Infect Dis* **192**, 1588-1596, doi:10.1086/496894 (2005).

229 **Riou, C. *et al.*** Distinct kinetics of Gag-specific CD4+ and CD8+ T cell responses during acute HIV-1 infection. *J Immunol* **188**, 2198-2206, doi:10.4049/jimmunol.1102813 (2012).

230 **Janssen, E. M. *et al.*** CD4+ T cells are required for secondary expansion and memory in CD8+ T lymphocytes. *Nature* **421**, 852-856, doi:10.1038/nature01441 (2003).

231 **Kamimura, D. & Bevan, M. J.** Naive CD8+ T cells differentiate into protective memory-like cells after IL-2 anti IL-2 complex treatment in vivo. *J Exp Med* **204**, 1803-1812, doi:10.1084/jem.20070543 (2007).

232 **Betts, M. R. *et al.*** Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection. *J Virol* **75**, 11983-11991, doi:10.1128/JVI.75.24.11983-11991.2001 (2001).

233 **Burton, D. R. & Mascola, J. R.** Antibody responses to envelope glycoproteins in HIV-1 infection. *Nat Immunol* **16**, 571-576, doi:10.1038/ni.3158 (2015).

234 **Tomaras, G. D. *et al.*** Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J Virol* **82**, 12449-12463, doi:10.1128/JVI.01708-08 (2008).

235 **Deeks, S. G. *et al.*** Neutralizing antibody responses against autologous and heterologous viruses in acute versus chronic human immunodeficiency virus (HIV) infection: evidence for a constraint on the ability of HIV to completely evade neutralizing antibody responses. *J Virol* **80**, 6155-6164, doi:10.1128/JVI.00093-06 (2006).

236 **Gray, E. S. *et al.*** The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *J Virol* **85**, 4828-4840, doi:10.1128/JVI.00198-11 (2011).

237 **Xiao, X. *et al.*** Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochem Biophys Res Commun* **390**, 404-409, doi:10.1016/j.bbrc.2009.09.029 (2009).

238 **Plotkin, S. A.** Immunologic correlates of protection induced by vaccination. *Pediatr Infect Dis J* **20**, 63-75 (2001).

239 **Karlsson Hedestam, G. B. *et al.*** The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat Rev Microbiol* **6**, 143-155, doi:10.1038/nrmicro1819 (2008).

240 **Prince, A. M. *et al.*** Prevention of HIV infection by passive immunization with HIV immunoglobulin. *AIDS Res Hum Retroviruses* **7**, 971-973, doi:10.1089/aid.1991.7.971 (1991).

241 **Mascola, J. R. *et al.*** Protection of Macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies. *J Virol* **73**, 4009-4018 (1999).

242 **Lifson, J. D. & Haigwood, N. L.** Lessons in nonhuman primate models for AIDS vaccine research: from minefields to milestones. *Cold Spring Harb Perspect Med* **2**, a007310, doi:10.1101/cshperspect.a007310 (2012).

243 **Hessell, A. J. *et al.*** Effective, low-titer antibody protection against low-dose repeated mucosal SHIV challenge in macaques. *Nat Med* **15**, 951-954, doi:10.1038/nm.1974 (2009).

244 **Rerks-Ngarm, S. *et al.*** Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med* **361**, 2209-2220, doi:10.1056/NEJMoa0908492 (2009).

245 **Haynes, B. F. *et al.*** Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med* **366**, 1275-1286, doi:10.1056/NEJMoa1113425 (2012).

246 **Mestecky, J., Raska, M., Novak, J., Alexander, R. C. & Moldoveanu, Z.** Antibody-mediated protection and the mucosal immune system of the genital tract: relevance to vaccine design. *J Reprod Immunol* **85**, 81-85, doi:10.1016/j.jri.2010.02.003 (2010).

247 **Goenka, R. *et al.*** Cutting edge: dendritic cell-restricted antigen presentation initiates the follicular helper T cell program but cannot complete ultimate effector differentiation. *J Immunol* **187**, 1091-1095, doi:10.4049/jimmunol.1100853 (2011).

248 **Choi, Y. S. *et al.*** ICOS receptor instructs T follicular helper cell versus effector cell differentiation via induction of the transcriptional repressor Bcl6. *Immunity* **34**, 932-946, doi:10.1016/j.immuni.2011.03.023 (2011).

249 **Choi, Y. S. *et al.*** Bcl6 expressing follicular helper CD4 T cells are fate committed early and have the capacity to form memory. *J Immunol* **190**, 4014-4026, doi:10.4049/jimmunol.1202963 (2013).

250 **Ballesteros-Tato, A. *et al.*** Interleukin-2 inhibits germinal center formation by limiting T follicular helper cell differentiation. *Immunity* **36**, 847-856, doi:10.1016/j.immuni.2012.02.012 (2012).

251 **Nurieva, R. I. *et al.*** Bcl6 mediates the development of T follicular helper cells. *Science* **325**, 1001-1005, doi:10.1126/science.1176676 (2009).

252 **Tubo, N. J. *et al.*** Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell* **153**, 785-796, doi:10.1016/j.cell.2013.04.007 (2013).

253 **Xu, H. *et al.*** Follicular T-helper cell recruitment governed by bystander B cells and ICOS-driven motility. *Nature* **496**, 523-527, doi:10.1038/nature12058 (2013).

254 **Crotty, S.** Follicular helper CD4 T cells (TFH). *Annu Rev Immunol* **29**, 621-663, doi:10.1146/annurev-immunol-031210-101400 (2011).

255 **Nurieva, R. I. & Chung, Y.** Understanding the development and function of T follicular helper cells. *Cell Mol Immunol* **7**, 190-197, doi:10.1038/cmi.2010.24 (2010).

256 **Johnston, R. J. *et al.*** Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325**, 1006-1010, doi:10.1126/science.1175870 (2009).

257 **Hatzi, K. *et al.*** BCL6 orchestrates Tfh cell differentiation via multiple distinct mechanisms. *J Exp Med* **212**, 539-553, doi:10.1084/jem.20141380 (2015).

258 **Hams, E. *et al.*** Blockade of B7-H1 (programmed death ligand 1) enhances humoral immunity by positively regulating the generation of T follicular helper cells. *J Immunol* **186**, 5648-5655, doi:10.4049/jimmunol.1003161 (2011).

259 **Schwickert, T. A. *et al.*** A dynamic T cell-limited checkpoint regulates affinity-dependent B cell entry into the germinal center. *J Exp Med* **208**, 1243-1252, doi:10.1084/jem.20102477 (2011).

260 **Good-Jacobson, K. L. *et al.*** PD-1 regulates germinal center B cell survival and the formation and affinity of long-lived plasma cells. *Nat Immunol* **11**, 535-542, doi:10.1038/ni.1877 (2010).

261 **Zhang, Y., Garcia-Ibanez, L. & Toellner, K. M.** Regulation of germinal center B-cell differentiation. *Immunol Rev* **270**, 8-19, doi:10.1111/imr.12396 (2016).

262 **Chung, Y. *et al.*** Follicular regulatory T cells expressing Foxp3 and Bcl-6 suppress germinal center reactions. *Nat Med* **17**, 983-988, doi:10.1038/nm.2426 (2011).

263 **Locci, M. *et al.*** Human circulating PD-1+CXCR3-CXCR5+ memory Tfh cells are highly functional and correlate with broadly neutralizing HIV antibody responses. *Immunity* **39**, 758-769, doi:10.1016/j.immuni.2013.08.031 (2013).

264 **Yamamoto, T. *et al.*** Quality and quantity of TFH cells are critical for broad antibody development in SHIVAD8 infection. *Sci Transl Med* **7**, 298ra120, doi:10.1126/scitranslmed.aab3964 (2015).

265 **Kohler, S. L. *et al.*** Germinal Center T Follicular Helper Cells Are Highly Permissive to HIV-1 and Alter Their Phenotype during Virus Replication. *J Immunol* **196**, 2711-2722, doi:10.4049/jimmunol.1502174 (2016).

266 **Brenchley, J. M. *et al.*** Differential infection patterns of CD4+ T cells and lymphoid tissue viral burden distinguish progressive and nonprogressive lentiviral infections. *Blood* **120**, 4172-4181, doi:10.1182/blood-2012-06-437608 (2012).

267 **Connick, E. *et al.*** CTL fail to accumulate at sites of HIV-1 replication in lymphoid tissue. *J Immunol* **178**, 6975-6983 (2007).

268 **Folkvord, J. M., Armon, C. & Connick, E.** Lymphoid follicles are sites of heightened human immunodeficiency virus type 1 (HIV-1) replication and reduced antiretroviral effector mechanisms. *AIDS Res Hum Retroviruses* **21**, 363-370, doi:10.1089/aid.2005.21.363 (2005).

269 **Perreau, M. *et al.*** Follicular helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and production. *J Exp Med* **210**, 143-156, doi:10.1084/jem.20121932 (2013).

270 **Connick, E. *et al.*** Compartmentalization of simian immunodeficiency virus replication within secondary lymphoid tissues of rhesus macaques is linked to disease stage and inversely related to localization of virus-specific CTL. *J Immunol* **193**, 5613-5625, doi:10.4049/jimmunol.1401161 (2014).

271 **Fukazawa, Y. *et al.*** B cell follicle sanctuary permits persistent productive simian immunodeficiency virus infection in elite controllers. *Nat Med* **21**, 132-139, doi:10.1038/nm.3781 (2015).

272 **Moukambi, F. *et al.*** Early Loss of Splenic Tfh Cells in SIV-Infected Rhesus Macaques. *PLoS Pathog* **11**, e1005287, doi:10.1371/journal.ppat.1005287 (2015).

273 **Yue, F. Y. *et al.*** HIV-specific IL-21 producing CD4+ T cells are induced in acute and chronic progressive HIV infection and are associated with relative viral control. *J Immunol* **185**, 498-506, doi:10.4049/jimmunol.0903915 (2010).

274 **Lindqvist, M. *et al.*** Expansion of HIV-specific T follicular helper cells in chronic HIV infection. *J Clin Invest* **122**, 3271-3280, doi:10.1172/JCI64314 (2012).

275 **Banga, R. *et al.*** PD-1(+) and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. *Nat Med* **22**, 754-761, doi:10.1038/nm.4113 (2016).

276 **Chowdhury, A. *et al.*** Decreased T Follicular Regulatory Cell/T Follicular Helper Cell (TFH) in Simian Immunodeficiency Virus-Infected Rhesus Macaques May Contribute to Accumulation of TFH in Chronic Infection. *J Immunol* **195**, 3237-3247, doi:10.4049/jimmunol.1402701 (2015).

277 **Kalisky, T., Blainey, P. & Quake, S. R.** Genomic analysis at the single-cell level. *Annu Rev Genet* **45**, 431-445, doi:10.1146/annurev-genet-102209-163607 (2011).

278 **Kalisky, T. & Quake, S. R.** Single-cell genomics. *Nat Methods* **8**, 311-314, doi:10.1038/nmeth0411-311 (2011).

279 **Raj, A. & van Oudenaarden, A.** Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-226, doi:10.1016/j.cell.2008.09.050 (2008).

280 **Raj, A. & van Oudenaarden, A.** Single-molecule approaches to stochastic gene expression. *Annu Rev Biophys* **38**, 255-270, doi:10.1146/annurev.biophys.37.032807.125928 (2009).

281 **Fukaya, T., Lim, B. & Levine, M.** Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358-368, doi:10.1016/j.cell.2016.05.025 (2016).

282 **Bengtsson, M., Stahlberg, A., Rorsman, P. & Kubista, M.** Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res* **15**, 1388-1392, doi:10.1101/gr.3820805 (2005).

283 **Brenchley, J. M. *et al.*** T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis. *J Virol* **78**, 1160-1168 (2004).

284 **Chun, T. W., Chadwick, K., Margolick, J. & Siliciano, R. F.** Differential susceptibility of naive and memory CD4+ T cells to the cytopathic effects of infection with human immunodeficiency virus type 1 strain LAI. *J. Virol.* **71**, 4436-4444 (1997).

285 **Abdel-Mohsen, M. *et al.*** Expression profile of host restriction factors in HIV-1 elite controllers. *Retrovirology* **10**, 106, doi:10.1186/1742-4690-10-106 (2013).

286 **Raposo, R. A. *et al.*** Dynamic regulation of host restriction factor expression over the course of HIV-1 infection in vivo. *J Virol* **88**, 11624-11629, doi:10.1128/JVI.01771-14 (2014).

287 **McLaren, P. J. *et al.*** Identification of potential HIV restriction factors by combining evolutionary genomic signatures with functional analyses. *Retrovirology* **12**, 41, doi:10.1186/s12977-015-0165-5 (2015).

288 **Tripathi, S. K. & Lahesmaa, R.** Transcriptional and epigenetic regulation of T-helper lineage specification. *Immunol Rev* **261**, 62-83, doi:10.1111/imr.12204 (2014).

289 **Lackner, A. A., Mohan, M. & Veazey, R. S.** The gastrointestinal tract and AIDS pathogenesis. *Gastroenterology* **136**, 1965-1978, doi:S0016-5085(09)00299-6 [pii] (2009).

290 **Rahmberg, A. R. *et al.*** Tetherin upregulation in simian immunodeficiency virus-infected macaques. *J Virol* **87**, 13917-13921, doi:10.1128/JVI.01757-13 (2013).

291 **Anonymous**. National Research Council of the National Academies. Guide for the Care and use of Laboratory Animals. *Washington, D.C. National Institutes of Health.* (2011).

292 **Abdel-Motal, U. M. *et al.*** Kinetics of expansion of SIV Gag-specific CD8+ T lymphocytes following challenge of vaccinated macaques. *Virology* **333**, 226-238, doi:10.1016/j.virol.2004.12.030 (2005).

293 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. (2014).

294 **Benjamini, Y. & Hochberg, Y.** Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).

295 **Storey, J. D.** A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498, doi:10.1111/1467-9868.00346 (2002).

296 **Ritchie, M. E. *et al.*** limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

297 **Pitcher, C. J. *et al.*** Development and homeostasis of T cell memory in rhesus macaque. *J. Immunol.* **168**, 29-43 (2002).

298 **Picker, L. J. *et al.*** IL-15 induces CD4 effector memory T cell production and tissue emigration in nonhuman primates. *J. Clin. Invest.* **116**, 1514-1524 (2006).

299 **Desimmie, B. A. *et al.*** Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J Mol Biol* **426**, 1220-1245, doi:10.1016/j.jmb.2013.10.033 (2014).

300 **Schmitt, K. *et al.*** Differential virus restriction patterns of rhesus macaque and human APOBEC3A: implications for lentivirus evolution. *Virology* **419**, 24-42, doi:10.1016/j.virol.2011.07.017 (2011).

301 **Taylor, H. E., Khatua, A. K. & Popik, W.** The innate immune factor apolipoprotein L1 restricts HIV-1 infection. *J Virol* **88**, 592-603, doi:10.1128/JVI.02828-13 (2014).

302 **Chen, H. *et al.*** CD4+ T cells from elite controllers resist HIV-1 infection by selective upregulation of p21. *J Clin Invest* **121**, 1549-1560, doi:10.1172/JCI44539 (2011).

303 **Bergamaschi, A. *et al.*** The CDK inhibitor p21Cip1/WAF1 is induced by FcgammaR activation and restricts the replication of human immunodeficiency virus type 1 and related primate lentiviruses in human macrophages. *J Virol* **83**, 12253-12265, doi:10.1128/JVI.01395-09 (2009).

304 **Garcia-Exposito, L. *et al.*** Gelsolin activity controls efficient early HIV-1 infection. *Retrovirology* **10**, 39, doi:10.1186/1742-4690-10-39 (2013).

305 **Woods, M. W. *et al.*** Human HERC5 restricts an early stage of HIV-1 assembly by a mechanism correlating with the ISGylation of Gag. *Retrovirology* **8**, 95, doi:10.1186/1742-4690-8-95 (2011).

306 **Burdick, R. *et al.*** P body-associated protein Mov10 inhibits HIV-1 replication at multiple stages. *J Virol* **84**, 10241-10253, doi:10.1128/JVI.00585-10 (2010).

307 **Furtak, V. *et al.*** Perturbation of the P-body component Mov10 inhibits HIV-1 infectivity. *PLoS One* **5**, e9081, doi:10.1371/journal.pone.0009081 (2010).

308 **Wang, X. *et al.*** Moloney leukemia virus 10 (MOV10) protein inhibits retrovirus replication. *J Biol Chem* **285**, 14346-14355, doi:10.1074/jbc.M110.109314 (2010).

309 **Bueno, M. T. *et al.*** Poly(ADP-ribose) polymerase 1 promotes transcriptional repression of integrated retroviruses. *J Virol* **87**, 2496-2507, doi:10.1128/JVI.01668-12 (2013).

310 **Boulanger, M. C. *et al.*** Methylation of Tat by PRMT6 regulates human immunodeficiency virus type 1 gene expression. *J Virol* **79**, 124-131, doi:10.1128/JVI.79.1.124-131.2005 (2005).

311 **Singhroy, D. N. *et al.*** Automethylation of protein arginine methyltransferase 6 (PRMT6) regulates its stability and its anti-HIV-1 activity. *Retrovirology* **10**, 73, doi:10.1186/1742-4690-10-73 (2013).

312 **Turelli, P. *et al.*** Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol Cell* **7**, 1245-1254 (2001).

313 **Singh, R. *et al.*** Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. *J Virol* **85**, 208-216, doi:10.1128/JVI.01810-10 (2011).

314 **Tissot, C. & Mechti, N.** Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus type 1 long terminal repeat expression. *J Biol Chem* **270**, 14891-14898 (1995).

315 **Fridell, R. A., Harding, L. S., Bogerd, H. P. & Cullen, B. R.** Identification of a novel human zinc finger protein that specifically interacts with the activation domain of lentiviral Tat proteins. *Virology* **209**, 347-357, doi:10.1006/viro.1995.1266 (1995).

316 **Zhang, F., Hatziioannou, T., Perez-Caballero, D., Derse, D. & Bieniasz, P. D.** Antiretroviral potential of human tripartite motif-5 and related proteins. *Virology* **353**, 396-409, doi:10.1016/j.virol.2006.05.035 (2006).

317 **Battivelli, E. *et al.*** Modulation of TRIM5alpha activity in human cells by alternatively spliced TRIM5 isoforms. *J Virol* **85**, 7828-7835, doi:10.1128/JVI.00648-11 (2011).

318 **Livak, K. J. & Schmittgen, T. D.** Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).

319 **Andersen, C. L., Jensen, J. L. & Orntoft, T. F.** Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* **64**, 5245-5250, doi:10.1158/0008-5472.CAN-04-0496 (2004).

320 **Raposo, R. A. *et al.*** Effects of cellular activation on anti-HIV-1 restriction factor expression profile in primary cells. *J Virol* **87**, 11924-11929, doi:10.1128/JVI.02128-13 (2013).

321 **Parlato, S. *et al.*** LOX-1 as a natural IFN-alpha-mediated signal for apoptotic cell uptake and antigen presentation in dendritic cells. *Blood* **115**, 1554-1563, doi:10.1182/blood-2009-07-234468 (2010).

322 **Henig, N. *et al.*** Interferon-beta induces distinct gene expression response patterns in human monocytes versus T cells. *PLoS One* **8**, e62366, doi:10.1371/journal.pone.0062366 (2013).

323 **Sanda, C. *et al.*** Differential gene induction by type I and type II interferons and their combination. *J Interferon Cytokine Res* **26**, 462-472, doi:10.1089/jir.2006.26.462 (2006).

324 **Rahm, N. & Telenti, A.** The role of tripartite motif family members in mediating susceptibility to HIV-1 infection. *Curr Opin HIV AIDS* **7**, 180-186, doi:10.1097/COH.0b013e32835048e1 (2012).

325 **Wang, Y. *et al.*** Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC Genomics* **7**, 59, doi:10.1186/1471-2164-7-59 (2006).

326 **Abdel-Mohsen, M. *et al.*** Effects of alpha interferon treatment on intrinsic anti-HIV-1 immunity in vivo. *J Virol* **88**, 763-767, doi:10.1128/JVI.02687-13 (2014).

327 **Xu, H., Wang, X., Morici, L. A., Pahar, B. & Veazey, R. S.** Early divergent host responses in SHIVsf162P3 and SIVmac251 infected macaques correlate with control of viremia. *PLoS One* **6**, e17965, doi:10.1371/journal.pone.0017965 (2011).

328 **Barouch, D. H. *et al.*** Rapid Inflammasome Activation following Mucosal SIV Infection of Rhesus Monkeys. *Cell* **165**, 656-667, doi:10.1016/j.cell.2016.03.021 (2016).

329 **Zhu, T. *et al.*** Evidence for human immunodeficiency virus type 1 replication in vivo in CD14(+) monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. *J Virol* **76**, 707-716 (2002).

330 **Fulcher, J. A. *et al.*** Compartmentalization of human immunodeficiency virus type 1 between blood monocytes and CD4+ T cells during infection. *J Virol* **78**, 7883-7893, doi:10.1128/JVI.78.15.7883-7893.2004 (2004).

331 **Li, S. *et al.*** Persistent CCR5 utilization and enhanced macrophage tropism by primary blood human immunodeficiency virus type 1 isolates from advanced stages of disease and comparison to tissue-derived isolates. *J Virol* **73**, 9741-9755 (1999).

332 **Uittenbogaart, C. H. *et al.*** Differential tropism of HIV-1 isolates for distinct thymocyte subsets in vitro. *AIDS* **10**, F9-16 (1996).

333 **Bakri, Y. *et al.*** The maturation of dendritic cells results in postintegration inhibition of HIV-1 replication. *J Immunol* **166**, 3780-3788 (2001).

334 **McNamara, L. A. *et al.*** CD133+ hematopoietic progenitor cells harbor HIV genomes in a subset of optimally treated people with long-term viral suppression. *J Infect Dis* **207**, 1807-1816, doi:10.1093/infdis/jit118 (2013).

335 **Carter, C. C. *et al.*** HIV-1 infects multipotent progenitor cells causing cell death and establishing latent cellular reservoirs. *Nat Med* **16**, 446-451, doi:10.1038/nm.2109 (2010).

336 **Williams, K. C. & Hickey, W. F.** Central nervous system damage, monocytes and macrophages, and neurological disorders in AIDS. *Annu Rev Neurosci* **25**, 537-562, doi:10.1146/annurev.neuro.25.112701.142822 (2002).

337 **He, J. *et al.*** CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature* **385**, 645-649, doi:10.1038/385645a0 (1997).

338 **Lin, Y. L. *et al.*** Cell surface CCR5 density determines the postentry efficiency of R5 HIV-1 infection. *Proc Natl Acad Sci U S A* **99**, 15590-15595, doi:10.1073/pnas.242134499 (2002).

339 **Kader, M., Bixler, S., Roederer, M., Veazey, R. & Mattapallil, J. J.** CD4 T cell subsets in the mucosa are CD28+Ki-67-HLA-DR-CD69+ but show differential infection based on alpha4beta7 receptor expression during acute SIV infection. *J Med Primatol* **38 Suppl 1**, 24-31, doi:10.1111/j.1600-0684.2009.00372.x (2009).

340 **Kader, M. *et al.*** Alpha4(+)beta7(hi)CD4(+) memory T cells harbor most Th-17 cells and are preferentially infected during acute SIV infection. *Mucosal Immunol* **2**, 439-449, doi:10.1038/mi.2009.90 (2009).

341 **Guo, J., Wang, W., Yu, D. & Wu, Y.** Spinoculation triggers dynamic actin and cofilin activity that facilitates HIV-1 infection of transformed and resting CD4 T cells. *J Virol* **85**, 9824-9833, doi:10.1128/JVI.05170-11 (2011).

342 **Monteiro, P. *et al.*** Memory CCR6+CD4+ T cells are preferential targets for productive HIV type 1 infection regardless of their expression of integrin beta7. *J Immunol* **186**, 4618-4630, doi:10.4049/jimmunol.1004151 (2011).

343 **Chang, S. T. *et al.*** Next-generation sequencing reveals HIV-1-mediated suppression of T cell activation and RNA processing and regulation of noncoding RNA expression in a CD4+ T cell line. *MBio* **2**, doi:10.1128/mBio.00134-11 (2011).

344 **Mehla, R. & Ayyavoo, V.** Gene array studies in HIV-1 infection. *Curr HIV/AIDS Rep* **9**, 34-43, doi:10.1007/s11904-011-0100-x (2012).

345 **Giri, M. S., Nebozhyn, M., Showe, L. & Montaner, L. J.** Microarray data on gene modulation by HIV-1 in immune cells: 2000-2006. *J Leukoc Biol* **80**, 1031-1043, doi:10.1189/jlb.0306157 (2006).

346 **Imbeault, M., Giguere, K., Ouellet, M. & Tremblay, M. J.** Exon level transcriptomic profiling of HIV-1-infected CD4(+) T cells reveals virus-induced genes and host environment favorable for viral replication. *PLoS Pathog* **8**, e1002861, doi:10.1371/journal.ppat.1002861 (2012).

347 **Mohammadi, P. *et al.*** 24 hours in the life of HIV-1 in a T cell line. *PLoS Pathog* **9**, e1003161, doi:10.1371/journal.ppat.1003161 (2013).

348 **Schultz, B. T. *et al.*** Circulating HIV-Specific Interleukin-21(+)CD4(+) T Cells Represent Peripheral Tfh Cells with Antigen-Dependent Helper Functions. *Immunity* **44**, 167-178, doi:10.1016/j.immuni.2015.12.011 (2016).

349 **Ma, C. S.** Human T Follicular Helper Cells in Primary Immunodeficiency: Quality Just as Important as Quantity. *J Clin Immunol* **36 Suppl 1**, 40-47, doi:10.1007/s10875-016-0257-6 (2016).

350 **Mylvaganam, G. H. *et al.*** Diminished viral control during simian immunodeficiency virus infection is associated with aberrant PD-1hi CD4 T cell enrichment in the lymphoid follicles of the rectal mucosa. *J Immunol* **193**, 4527-4536, doi:10.4049/jimmunol.1401222 (2014).

351 **Moguche, A. O. *et al.*** ICOS and Bcl6-dependent pathways maintain a CD4 T cell population with memory-like properties during tuberculosis. *J Exp Med* **212**, 715-728, doi:10.1084/jem.20141518 (2015).

352 **Allam, A. *et al.*** TFH cells accumulate in mucosal tissues of humanized-DRAG mice and are highly permissive to HIV-1. *Sci Rep* **5**, 10443, doi:10.1038/srep10443 (2015).

353 **Thomas, J. A., Ott, D. E. & Gorelick, R. J.** Efficiency of human immunodeficiency virus type 1 postentry infection processes: evidence against disproportionate numbers of defective virions. *J Virol* **81**, 4367-4370, doi:10.1128/JVI.02357-06 (2007).

354 **Hofmann-Lehmann, R. *et al.*** Sensitive and robust one-tube real-time reverse transcriptase-polymerase chain reaction to quantify SIV RNA load: comparison of one- versus two-enzyme systems. *AIDS Res Hum Retroviruses* **16**, 1247-1257, doi:10.1089/08892220050117014 (2000).

355 **Amara, R. R. *et al.*** Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. *Science* **292**, 69-74 (2001).

356 **Saeed, A. I. *et al.*** TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374-378 (2003).

357 **Llano, M. *et al.*** An essential role for LEDGF/p75 in HIV integration. *Science* **314**, 461-464, doi:10.1126/science.1132319 (2006).

358 **Subramanian, A. *et al.*** Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

359 **Novershtern, N. *et al.*** Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296-309, doi:10.1016/j.cell.2011.01.004 (2011).

360 **Smith, E. E. & Malik, H. S.** The apolipoprotein L family of programmed cell death and immunity genes rapidly evolved in primates at discrete sites of host-pathogen interactions. *Genome Res* **19**, 850-858, doi:10.1101/gr.085647.108 (2009).

361 **Wang, M., Windgassen, D. & Papoutsakis, E. T.** Comparative analysis of transcriptional profiling of CD3+, CD4+ and CD8+ T cells identifies novel immune response players in T-cell activation. *BMC Genomics* **9**, 225, doi:10.1186/1471-2164-9-225 (2008).

362 **Livak, K. J. *et al.*** Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. *Methods* **59**, 71-79, doi:10.1016/j.ymeth.2012.10.004 (2013).

363 **Dominguez, M. H. *et al.*** Highly multiplexed quantitation of gene expression on single cells. *J Immunol Methods* **391**, 133-145, doi:10.1016/j.jim.2013.03.002 (2013).

364 **Sanchez-Freire, V., Ebert, A. D., Kalisky, T., Quake, S. R. & Wu, J. C.** Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. *Nat Protoc* **7**, 829-838, doi:10.1038/nprot.2012.021 (2012).

365 **McDavid, A. *et al.*** Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461-467, doi:10.1093/bioinformatics/bts714 (2013).

366 **Schaerli, P. *et al.*** CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function. *J Exp Med* **192**, 1553-1562 (2000).

367 **Kim, C. H. *et al.*** Subspecialization of CXCR5+ T cells: B helper activity is focused in a germinal center-localized subset of CXCR5+ T cells. *J Exp Med* **193**, 1373-1381 (2001).

368 **Pavlidis, P. & Noble, W. S.** Analysis of strain and regional variation in gene expression in mouse brain. *Genome Biol* **2**, RESEARCH0042 (2001).

369 **Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A.** Regulation of noise in the expression of a single gene. *Nat Genet* **31**, 69-73, doi:10.1038/ng869 (2002).

370 **Guo, G. *et al.*** Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18**, 675-685, doi:10.1016/j.devcel.2010.02.012 (2010).

371 **Diehn, M. *et al.*** Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* **458**, 780-783, doi:10.1038/nature07733 (2009).

372 **Dalerba, P. *et al.*** Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* **29**, 1120-1127, doi:10.1038/nbt.2038 (2011).

373 **Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S.** Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600-1603, doi:10.1126/science.1119623 (2006).

374 **Newman, J. R. *et al.*** Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846, doi:10.1038/nature04785 (2006).

375 **Bar-Even, A. *et al.*** Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**, 636-643, doi:10.1038/ng1807 (2006).

376 **Ciuffi, A., Rato, S. & Telenti, A.** Single-Cell Genomics for Virology. *Viruses* **8**, doi:10.3390/v8050123 (2016).

377 **Sen, A. *et al.*** Innate immune response to homologous rotavirus infection in the small intestinal villous epithelium at single-cell resolution. *Proc Natl Acad Sci U S A* **109**, 20667-20672, doi:10.1073/pnas.1212188109 (2012).

378 **Akiba, H. *et al.*** The role of ICOS in the CXCR5+ follicular B helper T cell maintenance in vivo. *J Immunol* **175**, 2340-2348 (2005).

379 **Yu, D. *et al.*** The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity* **31**, 457-468, doi:10.1016/j.immuni.2009.07.002 (2009).

380 **Breitfeld, D. *et al.*** Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production. *J Exp Med* **192**, 1545-1552 (2000).

381 **Ueno, H.** Human Circulating T Follicular Helper Cell Subsets in Health and Disease. *J Clin Immunol* **36 Suppl 1**, 34-39, doi:10.1007/s10875-016-0268-3 (2016).

382 **Schmitt, N., Bentebibel, S. E. & Ueno, H.** Phenotype and functions of memory Tfh cells in human blood. *Trends Immunol* **35**, 436-442, doi:10.1016/j.it.2014.06.002 (2014).

383 **Haynes, N. M. *et al.*** Role of CXCR5 and CCR7 in follicular Th cell positioning and appearance of a programmed cell death gene-1high germinal center-associated subpopulation. *J Immunol* **179**, 5099-5108 (2007).

384 **Hannedouche, S. *et al.*** Oxysterols direct immune cell migration via EBI2. *Nature* **475**, 524-527, doi:10.1038/nature10280 (2011).

385 **Qi, H., Cannons, J. L., Klauschen, F., Schwartzberg, P. L. & Germain, R. N.** SAP-controlled T-B cell interactions underlie germinal centre formation. *Nature* **455**, 764-769, doi:10.1038/nature07345 (2008).

386 **Mowat, A. M. & Agace, W. W.** Regional specialization within the intestinal immune system. *Nat Rev Immunol* **14**, 667-685, doi:10.1038/nri3738 (2014).

387 **Hashiguchi, M. *et al.*** Naive CD4+ T cells of Peyer's patches produce more IL-6 than those of spleen in response to antigenic stimulation. *Immunol Lett* **141**, 109-115, doi:10.1016/j.imlet.2011.09.001 (2011).

388 **Seth, S. *et al.*** Abundance of follicular helper T cells in Peyer's patches is modulated by CD155. *Eur J Immunol* **39**, 3160-3170, doi:10.1002/eji.200939470 (2009).

389 **Mehandru, S. *et al.*** Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. *J Exp Med* **200**, 761-770, doi:10.1084/jem.20041196 (2004).

390 **Veazey, R. S. *et al.*** Dynamics of CCR5 expression by CD4+ T cells in lymphoid tissues during simian immunodeficiency virus infection. *J. Virol.* **74**, 11001-11007 (2000).

391 **Xu, H. *et al.*** Persistent Simian Immunodeficiency Virus Infection Drives Differentiation, Aberrant Accumulation, and Latent Infection of Germinal Center Follicular T Helper Cells. *J Virol* **90**, 1578-1587, doi:10.1128/JVI.02471-15 (2016).

392 **Xu, Y. *et al.*** Simian immunodeficiency virus infects follicular helper CD4 T cells in lymphoid tissues during pathogenic infection of pigtail macaques. *J Virol* **87**, 3760-3773, doi:10.1128/JVI.02497-12 (2013).

393 **Azizi, A., Ghunaim, H., Diaz-Mitoma, F. & Mestecky, J.** Mucosal HIV vaccines: a holy grail or a dud? *Vaccine* **28**, 4015-4026, doi:10.1016/j.vaccine.2010.04.018 (2010).

394 **Aamer, H. A., Rajakumar, P., Nyaundi, J. & Murphey-Corb, M.** Resistance to simian immunodeficiency virus low dose rectal challenge is associated with higher constitutive TRIM5alpha expression in PBMC. *Retrovirology* **11**, 39, doi:10.1186/1742-4690-11-39 (2014).

395 **Kamme, F. *et al.*** Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *J Neurosci* **23**, 3607-3615 (2003).

396 **Kurimoto, K. *et al.*** An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res* **34**, e42, doi:10.1093/nar/gkl050 (2006).

397 **Kurimoto, K., Yabuta, Y., Ohinata, Y. & Saitou, M.** Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat Protoc* **2**, 739-752, doi:10.1038/nprot.2007.79 (2007).

398 **Tang, F. *et al.*** mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377-382, doi:10.1038/nmeth.1315 (2009).

399 **Tang, F. *et al.*** Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6**, 468-478, doi:10.1016/j.stem.2010.03.015 (2010).

400 **Ramskold, D. *et al.*** Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**, 777-782, doi:10.1038/nbt.2282 (2012).

401 **Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R.** Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196, doi:10.1126/science.1245316 (2014).

402 **Lassen, K. G., Bailey, J. R. & Siliciano, R. F.** Analysis of human immunodeficiency virus type 1 transcriptional elongation in resting CD4+ T cells in vivo. *J Virol* **78**, 9105-9114, doi:10.1128/JVI.78.17.9105-9114.2004 (2004).

403 **Weaver, S. *et al.*** Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods* **50**, 271-276, doi:10.1016/j.ymeth.2010.01.003 (2010).

404 **Saliba, A. E., Westermann, A. J., Gorski, S. A. & Vogel, J.** Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res* **42**, 8845-8860, doi:10.1093/nar/gku555 (2014).

405 **Yoneyama, M. *et al.*** Direct triggering of the type I interferon system by virus infection: activation of a transcription factor complex containing IRF-3 and CBP/p300. *EMBO J* **17**, 1087-1095, doi:10.1093/emboj/17.4.1087 (1998).

406 **Collins, S. E., Noyce, R. S. & Mossman, K. L.** Innate cellular response to virus particle entry requires IRF3 but not virus replication. *J Virol* **78**, 1706-1717 (2004).

407 **Cornes, J. S.** Number, size, and distribution of Peyer's patches in the human small intestine: Part I The development of Peyer's patches. *Gut* **6**, 225-229 (1965).

408 **Lycke, N. Y. & Bemark, M.** The role of Peyer's patches in synchronizing gut IgA responses. *Front Immunol* **3**, 329, doi:10.3389/fimmu.2012.00329 (2012).

409 **Macpherson, A. J., McCoy, K. D., Johansen, F. E. & Brandtzaeg, P.** The immune geography of IgA induction and function. *Mucosal Immunol* **1**, 11-22, doi:10.1038/mi.2007.6 (2008).

410 **MacLennan, I. C. *et al.*** Extrafollicular antibody responses. *Immunol Rev* **194**, 8-18 (2003).

411 **Lee, S. K. *et al.*** B cell priming for extrafollicular antibody responses requires Bcl-6 expression by T cells. *J Exp Med* **208**, 1377-1388, doi:10.1084/jem.20102065 (2011).

412 **Xu, H., Wang, X., Malam, N., Lackner, A. A. & Veazey, R. S.** Persistent Simian Immunodeficiency Virus Infection Causes Ultimate Depletion of Follicular Th Cells in AIDS. *J Immunol* **195**, 4351-4357, doi:10.4049/jimmunol.1501273 (2015).

413 **Petrovas, C. *et al.*** CD4 T follicular helper cell dynamics during SIV infection. *J Clin Invest* **122**, 3281-3294, doi:10.1172/JCI63039 (2012).

414 **Miles, B. & Connick, E.** TFH in HIV Latency and as Sources of Replication-Competent Virus. *Trends Microbiol* **24**, 338-344, doi:10.1016/j.tim.2016.02.006 (2016).

415 **Belec, L. *et al.*** Cervicovaginal overproduction of specific IgG to human immunodeficiency virus (HIV) contrasts with normal or impaired IgA local response in HIV infection. *J Infect Dis* **172**, 691-697 (1995).

416 **Israel, Z. R. & Marx, P. A.** Nonclassical mucosal antibodies predominate in genital secretions of HIV-1 infected chimpanzees. *J Med Primatol* **24**, 53-60 (1995).

417 **Mestecky, J. *et al.*** Paucity of antigen-specific IgA responses in sera and external secretions of HIV-type 1-infected individuals. *AIDS Res Hum Retroviruses* **20**, 972-988, doi:10.1089/aid.2004.20.972 (2004).

418 **Schafer, F. *et al.*** Lack of simian immunodeficiency virus (SIV) specific IgA response in the intestine of SIV infected rhesus macaques. *Gut* **50**, 608-614 (2002).

419 **Mestecky, J., Moldoveanu, Z., Smith, P. D., Hel, Z. & Alexander, R. C.** Mucosal immunology of the genital and gastrointestinal tracts and HIV-1 infection. *J Reprod Immunol* **83**, 196-200, doi:10.1016/j.jri.2009.07.005 (2009).

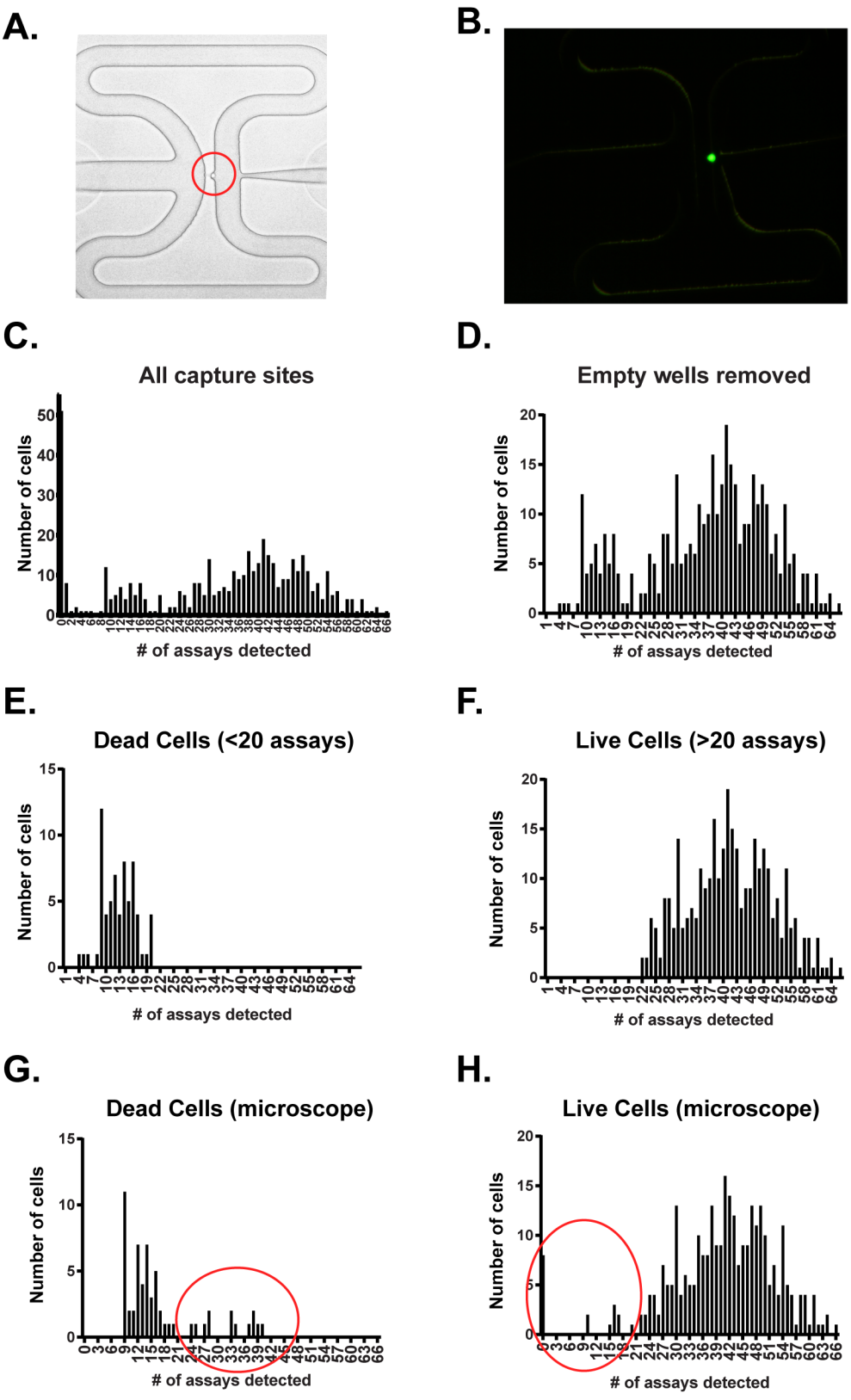
420 **Yukl, S. A. *et al.*** The Distribution of HIV DNA and RNA in Cell Subsets Differs in Gut and Blood of HIV-Positive Patients on ART: Implications for Viral Persistence. *J Infect Dis* **208**, 1212-1220, doi:10.1093/infdis/jit308 (2013).

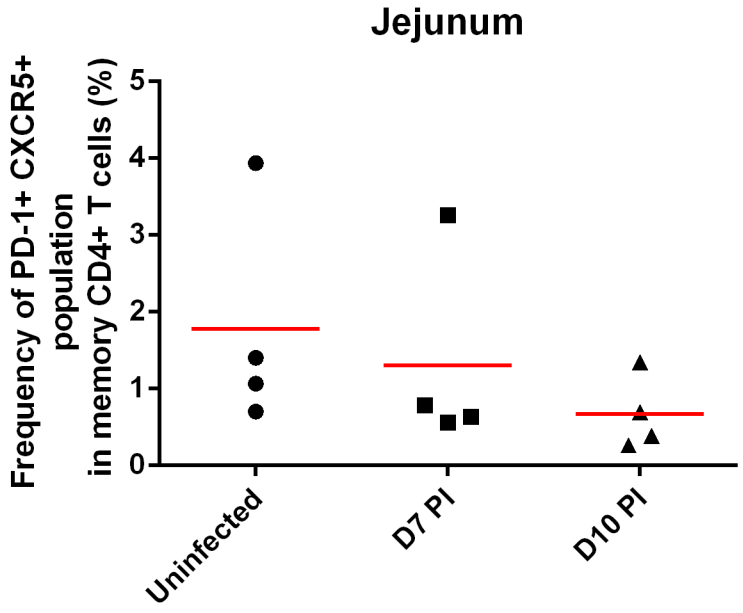
421 **Hatano, H. *et al.*** Comparison of HIV DNA and RNA in gut-associated lymphoid tissue of HIV-infected controllers and noncontrollers. *AIDS* **27**, 2255-2260, doi:10.1097/QAD.0b013e328362692f (2013).

Appendix A: Supplemental Figures

Supplemental Figure 3.1. Quality control procedure for single cell expression profiling. Visual inspection of each capture site allows tracking of the number of cells captured **(A)** and the live/dead staining status **(B)**. Histogram displaying the number of genes detectably expressed in all wells **(C)**. After removal of wells with fewer than 4 genes expressed, there are two clear populations of cells **(D)**, those that correlate with dead staining expressing fewer than 20 genes **(E)** and those that correlate with live staining expressing more than 20 genes **(F)**. Microscope staining alone allows some cells that stained dead (circled in **G**) and stained live (circled in **H**) to be incorrectly categorized from a transcriptional standpoint.

Supplemental Figure 3.1. (Continued)





Supplemental Figure 3.2. Frequency of PD-1+ CXCR5+ population in SIV infection. Changes were non-statistically significant (T-test).

Appendix B: Permissions

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Aug 30, 2016

This Agreement between Andrew Rahmberg ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	3938850123778
License date	Aug 30, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Cellular & Molecular Immunology
Licensed Content Title	Understanding the development and function of T follicular helper cells
Licensed Content Author	Roza I Nurieva and Yeonseok Chung
Licensed Content Date	Apr 12, 2010
Licensed Content Volume Number	7
Licensed Content Issue Number	3
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 2: Generation of Tfh cells.
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Gene Expression and CD4+ T Cell Susceptibility to SIV
Expected completion date	Sep 2016
Estimated size (number of pages)	170
Requestor Location	Andrew Rahmberg 954 Gatewood Road Room 3109 ATLANTA, GA 30329 United States

	Attn: Andrew Rahmberg
Billing Type	Invoice
Billing Address	Andrew Rahmberg 954 Gatewood Road Room 3109 ATLANTA, GA 30329 United States Attn: Andrew Rahmberg
Total	0.00 USD

Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer,* the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the

credit line should read:
Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer,* the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visithttp://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer,* the following credit line applies.
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.
Special Terms:
v1.1